DOCUMENT RESUME

ED 100 740 SO 008 020

AUTHOR	Larkins, A. Guy
TITLE	Critique of NAEP Objectives: Citizenship and Social Studies [and] Critique of NAEP Procedures--Task 1. Final Report Parts 1 and 2.
INSTITUTION	National Council for the Social Studies, Washington, D.C.
SPONS AGENCY	Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
PUB DATE	73
NOTE	115p.; Related documents are SO 008 019-026
EDRS PRICE	MF-$0.75 HC-$5.40 PLUS POSTAGE
DESCRIPTORS	*Citizenship; *Educational Assessment; *Educational Objectives; Educational Research; Evaluation; *Evaluation Criteria; Evaluation Methods; Measurement Goals; National Surveys; Research Criteria; *Social Studies
IDENTIFIERS	*National Assessment

ABSTRACT

The first part of Task 1 of the final report provides a critique of National Assessment of Educational Progress (NAEP) objectives. The report represents part of an investigation coordinated by a special Steering Committee working under the auspices of the National Council for the Social Studies. The first section examines whether the objectives meet the NAEP criteria that specialists in the subject area consider authentic from the viewpoint of the discipline, that school people recognize as desirable educational goals, and that parents agree are important for youth to know. The second section analyzes whether objectives meet the National Council for Social Studies Curriculum Guidelines of knowledge, abilities, valuing, and social participation. Section 3 examines whether the current objectives for citizenship and social studies overlap with objectives from other NAEP subject area assessments. Part 2 of the final report on Task 1 examines the criteria and procedures used to develop citizenship and social studies exercises for the assessment. The technical aspects of instrumentation, sampling, data analysis, and procedures are analyzed for validity and reliability. (DE)

ED100740

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CRITIQUE OF NAEP OBJECTIVES: CITIZENSHIP & SOCIAL STUDIES

TASK I

FINAL REPORT PART ONE

SO 008 020

A. Guy Larkins

University of Georgia

November, 1973

This report represents part of an investigation funded by the Education Commission of the States, and coordinated by a special Steering Committee working under the auspices of the National Council for the Social Studies.

CRITIQUE OF NAEP OBJECTIVES: CITIZENSHIP & SOCIAL STUDIES

Three Perspectives

How good are NAEP objectives in social studies and citizenship? This question can be answered from several perspectives, each posed as a separate question:

1. Do objectives in citizenship and social studies meet NAEP criteria?
2. Do they meet NCSS guidelines?
3. How much overlap is there between or among sets of objectives? (This question is based on two assumptions: that objectives should clearly differentiate among assessment areas, and that scarce resources should be protected by limiting unnecessary duplication.)

NAEP Criteria

It is reasonable to expect NAEP objectives in citizenship and social studies to meet criteria established by NAEP. These criteria are:

1. Specialists in the subject area must consider the objectives authentic from the viewpoint of the discipline. Scientists must agree the science objectives are authentic: mathematicians must agree upon the authenticity of the mathematics objectives, etc.
2. School people must recognize them as desirable goals for education and ones which schools are actively striving to achieve.
3. Parents and others interested in education must agree the objectives are important for youth and young adults to know, feel or understand (The National Assessment Approach to Exercise Development, page 12).*

---

*Most of the publications cited in this paper are not identifiable by author. Reference is generally made, therefore, to the title and page number. In some cases, titles have been abbreviated, but complete titles can be identified in the bibliography.

Academic Review

Some problems. The first criterion may reflect a time-honored assumption, that school subjects are based at least in part on parent academic disciplines. That assumption never completely resolves problems associated with what ought to be taught in schools and it seems to more easily fit some subjects than others. The examples given above, math and science, are among the more clear-cut instances of curricular areas which the assumption seems to fit. Unfortunately, citizenship is one of those which it does not. And the extent to which it is applicable to social studies is controversial.

Citizenship has no definitive parent discipline(s). A number of scholarly areas are relevant to what ought to be taught under citizenship, but no one discipline stands in the relation to citizenship that mathematics stands to school math. Nor can it be easily argued that any collection or combination of disciplines definitely stand for citizenship.

Part of the problem is obvious: there is no area of scholarly inquiry called citizenship. But, perhaps the central problem is not so apparent. More than other NAEP assessment areas, citizenship is concerned with beliefs, attitudes, and actions that require ethical justification. For instance, NAEP citizenship objectives are prefaced with statements such as: "Good 9-year-old citizens value others' right to. . ." or "Good adult citizens help anyone who is. . . (Citizenship Objectives, pages 12 and 14)." Not everyone agrees that citizenship objectives ought to have ethical directionality. Nevertheless, such language will probably strike most readers as

normal when applied to citizenship. But if math objectives were prefaced with "The good 17-year-old mathematician believes. . ." we would think the statement a bit odd. This is more than a semantic difference. It reflects the fact that the bulk of our concern with citizenship is centered in ethical values. The bulk of our concern with mathematics or chemistry is not. We can speak of educating a good chemist, just as we can of educating a good citizen, but the word "good" has an instrumental meaning in the first sense and an ethical meaning in the second. A chemist may pass judgement on what a student needs to know to understand chemistry or function well as a chemist. But what academician can tell us, as a natural outgrowth of his academic training, what a student needs to know to be a good citizen in the ethical sense of the term "good"?

This is not to deny that academic disciplines are relevant to assessing citizenship. But the relevance of the disciplines probably comes into play most strongly after major objectives are set, during the development of subobjectives or exercises. For instance, political science, considered only as an academic discipline, cannot tell us that the good citizen ought to respect the rights of others. But if knowing the Bill of Rights is determined on other grounds to be a desirable objective of citizenship, a political scientist or a constitutional historian can help determine the content validity of exercises designed to assess that objective. That is, they may know better than other academicians whether an objective faithfully reflects the Bill of Rights. They may even, as a result of their training, give more thought than

a laymen to issues of citizenship. Nevertheless, we cannot assume that citizenship can be adequately defined by modeling it after an academic discipline or even a set of disciplines.

Some people claim that the above comments do not apply to social studies. It is not uncommon for social studies to be defined as school subjects which parallel a fairly fixed set of academic disciplines: history, political science, sociology, anthropology, geography, and perhaps psychology or philosophy. Whether social studies ought to be defined that way is controversial. But that controversy aside, there are still problems. One of the most obvious is the apparent desirability of including in social studies content which is derived from no academic area. For instance, NAEP Social Studies Objective III (Are sensitive to creative-intuitive methods of explaining the human condition) and Objective V (Have a reasoned commitment to the values that sustain a free society) are not directly derivable from any of the social sciences or history.

As with citizenship, however, these comments should not be taken to mean that it is inappropriate to ask academics to pass judgement on the adequacy of objectives. They are meant simply as a caution that academic expertise is not necessarily always more relevant than lay opinion. To that extent, NAEP criteria are not equally applicable to each objective. NAEP was wise, therefore, to specify review of objectives by laymen and educators in addition to academic scholars.

Social studies objectives. Having commented on some limitations of using scholars to judge the validity of objectives, we now ask: How well was this criterion applied by NAEP?

The specific question we ask is, to what extent were representatives of the various social sciences included in the development of social studies objectives?

Frances S. Berdie claims:

> In the summer of 1965, 11 social scientists met for two and one-half days with members of the ETS staff to define the proper domain of an inquiry into the achievements of American education in this subject area (Social Studies Objectives, page 4).

Apparently, however, Berdie used the term "social scientist" loosely. Of the 11 members of that committee, only one was explicitly identified with an academic department within a university (Clyde F. Kohn, Department of Geography, University of Iowa). The others had various affiliations: four participants were affiliated with schools, two were affiliated with university departments of education, two were employed by social studies curriculum projects, one directed a teacher-training project and one directed a private institute.

At a later review conference, two persons clearly identified with academic disciplines were included. Appendix D of Social Studies Objectives (pages 31 and 32) lists nineteen participants in the Conference on Social Studies Objectives. Two participants are explicitly identified in NAEP publication with academic disciplines: David Easton, Professor of Political Science, University of Chicago and Lawrence Senesh, Professor of Economics, Purdue University. The extent to which Senesh is an active economist may be questioned. His major scholarly activity for the past several years appears to have been in curriculum development. He authored an elementary grades social studies program, Our Working World. The mimeographed report of the Conference on Social Studies

(Conference on Development, Appendix 2, page 2) identifies at least two other participants as speaking from the viewpoint of an academic discipline: Mr. Patterson, geography and Mrs. Collier, anthropology.

A third review conference is described by Berdie:

> . . .a committee of four social scientists (three of whom had served on the original ETS advisory committee) met to review the ETS staff's preliminary revision of the objectives (Social Studies Objectives, page 7).

Again, however, the word "social scientist" is apparently used in a broad and/or loose sense. All four participants on this committee were members of departments of education at their various universities.

In summary, few people who are clearly identifiable as specialists in academic disciplines reviewed social studies objectives. NAEP does not appear to have sought validity feedback from a broad spectrum of academic specialists in the social sciences and history.

Citizenship objectives. We now ask, "To what extent were representatives of the academic disciplines involved in the development of citizenship objectives?" Of six members of the Advisory Committee, two seem to fit this category: Joseph D. Lehman, Dean, School of Criminology, University of California, Berkely and Richard Lengaker, Chairman, Department of Political Science, University of California, Los Angeles. Of twenty-one persons listed as Teachers and Community Consultants, two seem to represent academic disciplines: Fred I. Greenstein, Associate Professor of Government, Wesleyan University and Nevitt Sanford, Director, Institute for Study of Human Problems (Citizenship Objectives, pages 47 - 50). Perhaps a third person could be added to this second list: Bernard C. Hennessy, Director, National

Center for Education in Politics. Some persons may object, of course, to including Lehman and Sanford as representatives of academic disciplines.

The above comments do not demonstrate that NAEP objectives in citizenship and social studies lack content validity to the academic disciplines. They do bear, however, on whether NAEP seriously attempted to satisfy the first of three criteria which it said must be met by all objectives. It appears that NAEP did not seek validity feedback from a broad spectrum of academic specialists in disciplines relevant to citizenship. Even if each relevant discipline was represented by one person, it would be difficult to argue that one person adequately represented the diversity of opinion within his specialty.

Taught in Schools

NAEP's second criterion for objectives is: School people must recognize them as desirable goals for education and ones which schools are actively striving to achieve. Do NAEP objectives in citizenship and social studies meet this criterion?

This second criterion is actually two: 1. Objectives must be considered desirable by school people; 2. Schools must be striving to achieve them. Logically, either criterion could be satisfied independently of the other. We assume that NAEP meant that both must be met; that each is a necessary criterion, but neither is sufficient.

Before offering evidence on whether NAEP citizenship and social studies objectives are goals which schools are actively striving to achieve, it is important to stress that this criterion has been consistently and repeatedly emphasized in NAEP reports.

It is listed as one of three criteria set by ". . .the governing Committee. . . for the development and acceptance of its objectives (What is National Assessment? page 5)." Similar statements appear in NAEP reports on objectives in each of the assessment areas. For instance, this criterion is mentioned at least four times in Citizenship Objectives (pages 2, 3, and 8) and at least an equal number of times in Social Studies Objectives (pages 2, 5, 7 and 8). A mimeographed report on NAEP conferences on the development of instrumentation contains numerous references to the importance of objectives reflecting actual school practice. One of the most emphatic of such statements is worth quoting:

> On December 15-16, a conference of the chairmen of the eleven panels from the four regional conferences of lay citizens was held in New York. Each chairman came to the meeting with prepared statements from his group which he presented at the meeting. . . . In response to comments on specific objectives, Chairman Tyler of the Exploratory Committee said he would take the suggestions back to the contractors for reconsideration by the teacher-scholar groups, but no objective would be added unless it was something the schools are earnestly seeking to attain. In other words, the assessment is to assess only what the schools are attempting to accomplish. (Conference on Development, page 4, italics added.)

Citizenship objectives. In light of the above statements, the following explanation by the contracting agency responsible for developing citizenship objectives is surprising:

> Our standard for choosing objectives was their importance to civic goals accepted by a consensus of our society. Whether the objective is being achieved well or poorly now, and how much schools, homes, churches, or other institutions have contributed to the present level of achievement, were not considered at all. We did NOT examine the current curricula of schools in the area of citizenship (Citizenship Objectives, page 5, italics in the original).

The inconsistency between NAEP stated criteria for objectives and the above quotation is not trivial. Repeated statements throughout NAEP publications create the expectation that NAEP is attempting to assess the outcomes of American schooling rather than educational outcomes of all social institutions taken together. If there is genuine confusion or disagreement over this point within NAEP or between NAEP and its contracting agencies, it ought to be cleared up. Objectives for assessing the outcomes of schooling are not identical to those for assessing the educational impact of the total society.

Despite the contracting agency's decision not to examine school curricula in citizenship, it is possible for their objectives to meet the stated criterion. It is this critic's opinion that the majority of NAEP citizenship objectives do meet that standard, but some do not.

It is doubtful that schools "are earnestly seeking to attain" the following objectives:

> IX. Help and Respect Their Own Families
>
> A. Respect the reasonable authority of their parents, or guardians, and help with home duties and problems.
>
> Age 17. . .They are willing to forego personal pleasures to help the family out of a crisis, and they typically accommodate their own plans to other family members' schedules (Citizenship Objectives, pages 43 44).
>
> IX. Nuture the Development of Their Children as Future Citizens (Adults)
>
> A. Provide for the basic needs and health of their children.
>
> They create a warm, accepting home environment. . . They always know where their children are and do not

leave young children without a baby-sitter (Citizenship Objectives, page 45).

It should be stressed that objectives developed to assess adult citizenship were apparently expected to meet the criterion of goals which schools earnestly seek to attain, at least at the time the adults were in school. At an NAEP directors meeting, "Chairman Tyler indicated that for adults, objectives should be in terms of what they should be expected to have obtained from a high school education (Conference on Development. . ., Appendix No. 3, page 2)."

The above two excerpts from Citizenship Objective IX are obviously atypical: They were deliberately chosen as extreme examples of failure to meet the criterion in question. Nevertheless, several citizenship objectives are tenuously related to school curricula, if they are related at all.

Social studies objectives. This author has claimed that at least some NAEP citizenship objectives fall outside the usual domain of schools. Is the same claim true of NAEP social studies objectives?

The answer appears to be "No," but only if social studies objectives are considered one at a time. When social studies objectives are viewed as a total set, the verdict is reversed. If we take the total set of social studies objectives as even approximately descriptive of school social studies curricula, children would receive an incredible dose of history and the social sciences. The following is a much abbreviated list of

what thirteen-year-old children would be expected to know:

> the distinctive knowledge domains of the social sciences
>
> the American economy in its international setting, including the role of world trade
>
> major similarities and differences between American and other economic systems
>
> major characteristics of life in Latin America, Europe, Africa, and Asia
>
> major events and leaders in United States history
>
> the traditional Indian cultures, the Spanish conquests, and national independence movements in Latin America
>
> significant developments in the early history of Western Europe
>
> significant historical developments in Africa and the Middle East
>
> major developments in Asian history
>
> characteristics of the major systems of government (Social Studies Objectives, pages 13-25).

Someone has got to be kidding. Does NAEP seriously believe that, by the time children are thirteen, schools have earnestly tried to give instruction in major concepts from each of the social sciences plus significant events in the history of the entire world? Should schools attempt to provide that much social studies instruction during the first six or seven grades?

The above listing of expected information is contained in only one of five social studies objectives. In addition, students are expected to learn "analytic-scientific procedures," "creative-intuitive methods" and "have a reasoned commitment to the values that sustain a free society (Social Studies Objectives, pages 9-12, 26-67)."

If this criticism seems too harsh, the reader may want to consider part of Professor Tyler's justification for the three NAEP criteria for objectives: "This careful attention to the identification of objectives should help to minimize the criticism frequently encountered with current tests in which some item is. . . crit cized by school people as something not in the curriculum (Citizenship Objectives, page 3)." Certainly, if it is reasonable to expect that the content of individual items be something that is taught in the curriculum, it is also reasonable to expect a set of objectives to approximate cumulative social studies content at a given grade or age-level.

A difficulty, which NAEP may have faced in producing objectives, is nearly inherent in the criteria used. It would be noteworthy for scholars in the academic disciplines, professors of education, school people and laymen to reach uninimity about what should be taught to children. This observation should not be taken to mean that NAEP used the wrong criteria. It should rather temper our criticism: NAEP should not be expected either to do the impossible or to take the easy road of dropping good criteria that cannot be met perfectly.

Lay Review

Possible results of inherent conflict among NAEP criteria can be illustrated by considering the third criterion, which is:

> Parents and others [laymen] interested in education must agree that objectives are important for youth and young adults to know, feel or understand (The National Assessment Approach to Exercise Development, page 12).

Apparently, serious attempts to meet this criterion in social studies were at least partly responsible for failure to produce objectives which reasonably approximate the content of school curricula. During late November and early December, 1965, eleven conferences were held at four locations, each in a different major region of the United States. At these conferences a total of 99 laymen from 34 states reviewed objectives in each of the NAEP assessment areas. On December 15-16, chairmen of the eleven lay panels met in New York to summarize the recommendations of their groups. At that meeting:

> About half of the chairmen reported that their groups felt the Social Studies section should be redefined and reworked so that the objectives would assess the degree to which there is basic subject competence in such fields as history, geography, economics, political science, sociology, and so forth (Conference on Development, Appendix #1, page 7).

This recommendation led to revision of the social studies objectives; greater emphasis was given to major ideas from history and the social sciences. Perhaps, structuring the total set of objectives to adequately reflect major content of the academic disciplines made it difficult for the same set of objectives to realistically approximate social studies curricula in the schools. NAEP may have been caught between the horns of what laymen thought should be assessed and what educators knew schools are attempting to teach.

How well was the third criterion met? Compared to usual practice, it appears that NAEP made an earnest effort to meet it. Despite the assumption that American public schools are answerable to citizens, assessments of educational achievement generally do

not involve laymen in determining what ought to be measured. That NAEP held a series of regional lay conferences to review objectives, and that revisions resulted from those conferences, indicates the third criterion was not taken lightly.

Could NAEP have done better? Yes, in at least two ways, one of which would have been expensive. First, NAEP could have improved the diversity of laymen involved. Selection procedures nearly guaranteed an over-representation of middle-class America. Citizens who attended the lay conferences were nominated by groups such as the American Federation of Labor, National P.T.A., the United States Chamber of Commerce and the National Association for the Advancement of Colored People. Persons who are genuinely outside the mainstream of American society are seldom both visible and acceptable to national organizations of this sort. It is difficult to defend excluding such persons from an operational definition of "concerned" or "intelligent" citizens (terms used repeatedly in NAEP descriptions of the lay panels). Selection of "intelligent" and "concerned" citizens who are not part of the visible mainstream of society need not have appreciably increased NAEP costs.

Second, if it were possible to ignore cost, NAEP could have used sample survey procedures to select a lay review panel representative of the general public. This approach would have produced greater diversity of representation, and it would have increased the probability that laymen in general would agree that the objectives are worth assessing. It should be restressed, however, that sample surveys are expensive; the

the benefits might not be worth the cost.

A third consideration is the careful use of key terms. NAEP publications misuse the term "representative" when describing lay panels. For instance:

> The conferees represented large city, suburban and rural or small town school systems in the four regions of the United States.
>
> There was also wide representation from lay organizations interested in education from all levels of activity: national, state and local (Conference on Development, page 11).

A more accurate claim would be that conferees were from large cities, suburbs, etc. The sample of conferees was hardly representative of those populations in the sense that researchers speak of representativeness. The loose use of "representative" when speaking of lay panels stands in stark contrast to the carefully qualified use of the term when describing samples of children, youth and young adults who were selected to participate in National Assessment.

Summary

The above sections dealt with the first of three perspectives from which we can ask: How good are NAEP objectives in social studies and citizenship? Phrased as a question, the first perspective was: Do objectives in citizenship and social studies meet NAEP criteria? The first criterion involved review by scholars in the academic disciplines. The best summary of how well NAEP met that criterion is: Few people who are clearly identifiable as scholars in relevant academic disciplines were listed as reviewers of social studies and citizenship objectives.

The second criterion required that objectives be considered desirable by school people and that they be "ones which schools are actively striving to achieve." The best summary statements of how well NAEP met this criterion appears to be: (1) Some NAEP citizenship objectives fall outside the usual domain of schools; (2) Taken as a whole, social studies objectives imply far more content than schools seriously attempt to teach, especially in the first seven grades.

The third criterion required lay review of all objectives. This criterion was taken seriously and achieved reasonably well. Greater diversity of social-class representation on lay review panels, however, may be desirable.

The following section deals with the second perspective: How well do NAEP objectives meet guidelines of the National Council for the Social Studies? The reader may notice that the second perspective amounts to another set of criteria for judging NAEP objectives. Those criteria are substantially different than some of the ones considered above. The result of this shift in criteria is that NAEP may be praised in the following section for characteristics which were considered negative when judged by the preceding standards, and vice versa. The obvious moral is that readers should not accept the various sets of criteria at face value, but decide which ones they consider most appropriate.

NAEP Objectives and NCSS Guidelines

Preface

Two position papers--"Standards for Social Studies Teachers" and "Social Studies Curriculum Guidelines"--were published by the National Council for the Social Studies in the December, 1971, issue of Social Education. Some may question whether it is reasonable to expect National Assessment to meet standards expressed in those papers; NCSS standards and guidelines were published after NAEP objectives were developed. Educators who take seriously both the NCSS guidelines and NAEP objectives, however, need some estimate of the fit between them.

Of the two NCSS papers, the second is most relevant to National Assessment. The first--"Standards for Social Studies Teachers"--deals primarily with selection and training of teachers and with some of their professional duties and appropriate behaviors. Objectives for students are seldom dealt with and then only by implication. Detailed comparison of NCSS standards for teachers with NAEP objectives is, therefore, not possible.

Major portions of the second paper--"Social Studies Curriculum Guidelines"--can be compared to NAEP objectives, but not easily. One problem is that the guidelines, and the rationale which prefaces them are vague. They most often deal in generalities without explicit examples. For instance, much of the rhetoric in the paper on guidelines indicates to this author that NCSS holds that

rational political citizenship is the major goal of social studies, but the guidelines do not use the term "citizenship."

A second problem in comparing NCSS guidelines to NAEP objectives is that the latter focus on outcome behavior; that is, on behavior which is supposedly indicative of good citizenship. NCSS guidelines sometimes mention outcome behavior, but the central focus is on instructional input. A straightforward comparison of output statements to input statements is not always possible, and some of the input guidelines are more easily translated to student outcomes than are others. As a rule-of-thumb, NCSS guidelines 1.0 to 5.0 are relevant to student behavior and are, therefore, more comparable to NAEP objectives than are guidelines 6.0 to 9.0. The NCSS position paper on curriculum guidelines, however, contains a section on rationale in addition to the listing of guidelines. Most of the rationale can be usefully compared to NAEP objectives.

A third difficulty in comparing guidelines to objectives is that there are two statements of NAEP positions--the assessments titled, "citizenship" and "social studies." Those statements differ on several major points. Care must be taken, therefore, to avoid implying that NAEP has a single, clear-cut position on each of the issues discussed below. In the discussion which follows, reference will be made to either "the position implied by NAEP social studies objectives" or "the position implied by NAEP citizenship objectives" rather than "the NAEP position."

Despite some vagueness in the NCSS paper on curriculum guidelines, lack of direct parallels between the guidelines and NAEP objectives, and inconsistency between NAEP citizenship and social studies objectives, each of the three documents contain points which are well worth consideration by social studies educators.

The structure of most of what follows is taken from the NCSS rationale, which is divided into four sections: Knowledge, Abilities, Valuing, and Social Participation. The NCSS position on each of these topics will be compared in turn to the positions expressed in NAEP citizenship and social studies objectives.

Knowledge

NCSS guidelines and NAEP social studies objectives. NCSS guidelines and NAEP social studies objectives agree that history and the social sciences are essential sources of knowledge for social studies. For instance, NAEP devotes more space to Objective IV, which states which content from history and the social sciences should be included in social studies, than to Objectives I, II, III, and V combined (Social Studies Objectives, pages 9-27). And NCSS proclaims:

> The traditional and obvious sources of knowledge for social studies are the social science disciplines. We include history here, of course, under the social science rubric (Social Education, December, 1971, page 856).

Despite agreement that history and the social sciences are essential sources of knowledge for social studies, the NCSS position paper and NAEP social studies objectives represent different views of how social studies content should be selected.

NAEP social studies objectives are prefaced with a definition, part of which states:

> Social studies is. . .a shorthand term for such subjects as history, geography, economics, political science, anthropology, sociology, and social psychology (Social Studies Objectives, page 9).

When read in context, the above quotation seems to be a descriptive definition of social studies as it is generally taught in American schools. It also appears, however, to be a prescriptive definition to the extent that thirteen of eighteen pages of NAEP social studies subobjectives are congruent with the view that social studies is a shorthand term for history and the social sciences (ibid, pages 13-25).

In contrast, NCSS explicitly rejects the ". . .arbitrary and limiting assumption that social studies and the social sciences are identical (Social Education, page 857)."

The NCSS and NAEP views on social studies are not as dichotomous as the above quotations suggest. NAEP does not limit social studies to history and the social sciences (see Objectives I, III, and IV in Social Studies Objectives). And, as we have seen, NCSS does not exclude history and the social sciences from its view of social studies. There is, nevertheless, a crucial difference concerning the role of history and the social sciences in the views expressed by the NCSS guidelines and NAEP social studies objectives.

A crucial difference is that the NAEP social studies assessment seems to assume that those portions of the curriculum which are based on history and the social science should be condensed replicas

of the parent disciplines. In contrast, the NCSS position seems to be that those portions of the curriculum which are selected from history and the social sciences should be selected for their relevance to the resolution of social problems.

The following are examples of how the NAEP social studies assessment treats objectives on history and the social sciences:

> IV. Have Knowledge Relevant to the Major Ideas and Concerns of Social Scientists
>
> A. Understand some of the distinctive modes of inquiry (questions and approaches) of social scientists.
>
> (1) Are aware of the distinctive knowledge domains of the social sciences, for example, that:
>
> Historians seek to record human events and to interpret them in a way that gives meaning to continuity and change;
>
> Anthropologists are interested in cultures and in the effects of a culture on the life of an individual;
>
> . . . .
>
> (2) Recognize the types of questions that are of major significance to several of the social sciences, such as:
>
> What explains the spatial patterns of phenomena on the face of the earth?
>
> How is the power of leaders structured by the government (Social Studies Objectives, page 13)?

In contrast, the following seems to represent the NCSS position on the relation of history and the social sciences to social studies:

> The notion that the disciplines must always be studied in their pure form. . .is insufficient for a curriculum intended to demonstrate the relationship between knowledge and rationally based social participation. It is true that the social sciences can make marked contribution to clarifying the basic issues which continue to require social attention. But the efforts of social scientists to develop an understanding of human behavior through research

> are not necessarily related to society's persistent problems and are seldom intended to arrive at the resolution of value conflicts or the formulation of public policy (Social Education, page 875).

As the above quotation indicates, the importance of the dispute over the proper role of history and the social sciences in the social studies curriculum is largely determined by one's view of the relation of social studies to citizenship education. NCSS appears to define social studies primarily, but not entirely, as citizenship education. The NAEP social studies assessment appears to define social studies primarily, but not entirely, as a condensed replica of history and the social sciences. Evidence has been offered for the latter claim. The following is offered as evidence that NCSS defines social studies as citizenship education.

That political citizenship is a major goal of social studies as advocated by NCSS is strongly implied by repeated use of the rhetoric of a public issues rationale in the NCSS position paper. That rationale, as stated elsewhere, equates social studies with education for political citizenship. Elements of a public issues rationale appear in the above quotation in which NCSS implies that social studies should lead to "rationally based social participation," should be "related to society's persistent problems," and should be relevant to "the formation of public policy." Other elements of a public issues rationale which appear in the NCSS position paper include: reference to human dignity as an ultimate goal which is enhanced by rational processes, explicit mention of

Constitutional values as enhancing human dignity; a disclaimer that social science content is automatically valid content for social studies; mention of the need to teach students to recognize and deal with factual, definitional, and value disputes; and frequent reference to the importance of preparing students to deal with social issues. The NCSS rationale for curriculum guidelines also claims that social problems "constitute the major concern of the social studies curriculum (Social Education, page 857, italics added.)" Such statements help explain the NCSS view that relevance to social problems is an important criterion for the selection of content from history and the social sciences, and why NCSS rejects the view that history and the social sciences should always be taught "in their pure form."

A word of caution, however, is in order. As previously stated, the NAEP social studies objectives appear to view social studies as condensed replicas of history and the social sciences, but they also reflect concern for political citizenship. Objective V, for instance, is: Have a reasoned commitment to the values that sustain a free society (Social Studies Objectives, page 26). The point of dispute between NCSS and the NAEP social studies assessment is not over whether citizenship education is important, but over the identification of appropriate criteria for selecting content from history and the social sciences. Although relevance to citizenship education is one criterion strongly implied by NCSS, and even though NCSS has clearly rejected some criteria, there is substantial lack of clarity within the NCSS position on this point. Guideline 3.8, for instance, is

tantalizingly vague:

> The program must include a careful selection from the disciplines of that knowledge which is of most worth (Social Education, page 862).

Disagreement and lack of clarity over appropriate criteria for selecting content from history and the social sciences extend beyond NCSS and NAEP to the profession as a whole. In weighing the merits of NAEP social studies objectives and NCSS curriculum guidelines, social studies educators should consider the clarity of their own professional commitments concerning the proper role of the academic disciplines.

An additional point of agreement between NCSS and NAEP needs to be noted. NCSS explicitly states that knowledge from sources other than history and the social sciences is appropriate. The following NAEP social studies objectives are consistent with that position:

> III. Are Sensitive to Creative-Intuitive Methods Of Explaining the Human Condition
>
> B. Obtain insight into human affairs from history and philosophy, and from fiction and other forms of art (Social Studies Objectives, page 12).

NAEP citizenship objectives. Despite vagueness in the NCSS position paper on curriculum guidelines, arguments have been offered that NCSS views citizenship education as a major function of social studies. And, despite rejection of the assumption that social studies are condensed replicas of the academic disciplines, NCSS asserts that "there could be no social studies without the social sciences (Social Education, page 857)." From the NCSS viewpoint, then, it is reasonable to ask:

To what extent are NAEP citizenship objectives based on appropriate content from history and the social sciences?

Most of the contributions by academic disciplines to NAEP citizenship objectives come from political science and history. Contributions also come from other fields, such as sociology and law. The manner in which NAEP apparently drew upon these disciplines is generally consistent with the approach advocated by NCSS. Judicious selection from the diciplines, rather than replication of them, seems to have been the rule. The following sample is typical of those NAEP citizenship objectives which require that knowledge from the social sciences and history be applied to social problems:

> I.F. They know of the vicious circle connecting lack of education, unemployment, poverty, and slums. They understand that these conditions contribute to apathy, alienation, lack of initiative, and crime (Citizenship Objectives, page 13).
>
> I.G. They know of the existence, both near home and over most of the world, of poverty, illiteracy, and disease (ibid., page 14).
>
> II. . . .know the rights guaranteed by the Constitution. . . They know that the law assumes a person to be innocent unless proven guilty. They know that the U. S. Constitution and the Bill of Rights protect these basic rights for all Americans (ibid., pages 14-15).
>
> IV. Know the main structure and functions of our governments (ibid., pages 21-26).
>
> V. Understand problems of international relations (ibid., page 31).

Although NAEP citizenship objectives are generally consistent with the NCSS position on the relation of history and the social sciences to the curriculum, there are important differences concerning the scope of citizenship education. Those differences have

implications for knowledge which should be included in social studies/ citizenship. Several NAEP citizenship objectives have little apparent relevance to the major thrust of NCSS guidelines.

We have considered some evidence that NCSS stresses political citizenship. Typical of that emphasis is the strong implication that social participation is an ultimate goal of social studies. Social participation is defined by NCSS as ". . .individual behavior guided by the values of human dignity and rationality and directed toward the resolution of problems confronting society (Social Education, page 859)."

A similar view of citizenship can be found in many NAEP citizenship objectives. For example, the objectives quoted above fit that view. Other NAEP citizenship objectives, however, range from health and safety (Citizenship Objectives, page 11), to family relations (ibid., pages 43-46) to social etiquette (ibid., page 10) to dating (loc. cit.) to the Boy Scout Motto (ibid., page 12 and 18) to vocational education (ibid., pages 41-42). That NAEP takes a broader view of citizenship than does NCSS should not, however, pose a serious problem for social studies educators unless they unthinkingly assume that social studies is responsible for education in all types of citizenship assessed by NAEP.

Abilities

NCSS guidelines and NAEP social studies objectives. NCSS defines abilities as including ". . .intellectual, data processing, and human relations competencies (Social Education, page 857)."

Among intellectual competencies, NCSS stresses divergent thinking and valuing. Inasmuch as valuing is the topic of a separate division of the NCSS rationale, it will be dealt with in a later section of this paper.

NCSS defines data processing as ". . .competence to locate and compile information, to present and interpret data, and to organize and assess source material (ibid., page 858)." Special mention is made of the need for ". . .higher levels of proficiency in data processing skills--for example, identifying hypotheses, making warranted inferences, and reading critically. . .(loc cit.)."

Among human relations skills, NCSS explicitly mentions ". . . a sensitivity to the needs of others, adequately developed communication skills, and the ability to cope with conflict and authority (loc. cit.)."

How close is the fit between NAEP social studies objectives and the intellectual skills cited by NCSS? The most explicit references among NCSS social studies objectives that bear upon divergent thinking are:

III. Are sensitive to creative-intuitive methods of explaining the human condition.

C. Recognize the role of creative-intuitive methods in scientific inquiry.

Ages 17, Adult — Recognize that creativity is an important element in a scientist's formulation of questions and hypotheses and in his efforts to synthesize aspects of human knowledge (Social Studies Objectives, page 12).

The wording of these objectives does not point to the need to foster or assess divergent thinking among students, but rather to assess whether they know that divergent thinking is useful to scientists. This emphasis does not meet the spirit of NCSS comments on creativity-divergent thinking.

Although NAEP social studies objectives do not emphasize divergent thinking, they are consistent with data processing competencies stressed by NCSS. The following NAEP social studies objectives illustrate the full range of data processing skills cited by NCSS and also illustrate, incidentally, the obvious overlap between intellectual competencies, data processing skills and valuing.

II. Use analytic-scientific procedures effectively.

- A. Identify and define problems and issues.
- B. Formulate generalizations and hypotheses capable of being tested.
- C. Obtain information from a variety of sources.
- D. Distinguish facts from opinions, relevant from irrevelant information, and reliable from unreliable sources.
- E. Detect logical errors, unstated assumptions, and unwarranted assertions; question unsupported generalizations; are aware of the complex nature of social causation and understand that sequence or relationship does not necessarily imply causation.
- F. Use data and evaluative criteria to make decisions (ibid., pages 10-11).

Of the three human relations competencies cited by NCSS--sensitivity to others, communication skills, and ability to cope with conflict and authority--the first and part of the third are mentioned in NAEP social studies objectives:

I.B. Are open to new information and ideas.

Ages 17, Adult — Obtain information and ideas from sources likely to have conflicting viewpoints.

I.C. Try to understand why other people think and act as they do.

Ages 17, Adult — Attempt to understand the views of others; are willing to put themselves into positions of others; share the feelings of others (ibid., pages 9-10).

Communication skills and ability to cope with authority are generally neglected in NAEP social studies objectives. Some people, however, might read Objective IV.F., which deals with characteristics of American government, as relevant to coping with authority.

In summary, the fit between NCSS and NAEP social studies objectives is tightest on data processing skills. The fit on divergent thinking and human relations skills is much looser.

NAEP citizenship objectives. Although numerous intellectual data processing, and human relations competencies are included in NAEP citizenship objectives, divergent-creative thinking is neglected. Furthermore, the general tone of many objectives is toward conformity and convergent thinking. Examples of conformity will be given in the discussion of valuing. Significant exceptions to the press for conformity are included immediately below in the discussion of human relations skills.

Data processing and communication are stressed in the following citizenship objectives. As previously noted, NCSS considers human relations skills to be part of communication competencies.

VII. Support rationality in communication, thought and action on social problems.

A. Try to inform themselves on socially important matters and to understand alternative viewpoints.

B. Evaluate communications critically and form their own opinions independently.

. . .

E. Support free communication and communicate honestly with others (Citizenship Objectives, pages 35-38).

The full positive import of citizenship objectives can only be appreciated by considering age-level examples provided by NAEP. The following objective includes an age-level example which is relevant to communication skills.

V.E. Apply democratic procedures on a practical level when working in a group.

> Age 17 . . .They support the right of dissenting views to be voiced and encourage adequate discussion before voting. They abide by democratically determined decisions but know the established procedures for trying to change a decision (persuasion, argument, petition, etc.). They mediate, and seek compromise and common ground when others disagree. They are willing to give in when the situation calls for some immediate action or when their objection is relatively unimportant. They understand the responsibilities involved in accepting leadership (e.g., to keep informed on relevant matters; to clarify issues, sum up discussion, and present suggestions to the group . . .(ibid., pages 30-31).

NAEP social studies objectives seem to be stronger than NAEP citizenship objectives on data processing skills, particularly those competencies which are often associated with critical thinking. But NAEP citizenship objectives are stronger than social studies objectives in human relations skills. Neither assessment emphasizes divergent-creative thinking. Social studies educators who take seriously the NCSS rationale and guidelines on abilities

should pick and choose among NAEP social studies and citizenship assessments for the most appropriate objectives.

Valuing

NCSS guidelines and NAEP social studies objectives. NCSS asserts that schools should not evade questions of value. In fact, statements such as the following imply that such questions are central to social studies:

> The primary purpose of a social studies program is neither to advance the frontiers of knowledge nor to produce social scientists. Rather its task is to engage students in analyzing and attempting to resolve the social issues confronting them (Social Education, page 861).

NCSS opposes blind indoctrination of values, but is ambiguous over whether inculation of certain values is justified. For instance, should schools expect students to know, prefer and act in accordance with core values such as those contained in the Bill of Rights? The following expresses that ambiguity:

> Neither young people nor society will deal constructively with present social realities through blind acceptance of specific ways of behaving, or of particular positions on public issues, or even of basic cultural values. . . .
>
> Still perplexing is the role of the school as an agent for inculating in the young widely held societal norms, standards of behavior, and ideological preferences. The issue is clouded with conflicting attitudes held by various groups. Cultural pluralism in America rightly hinders the school from seeking or producing uniform values among its students (ibid., page 859).

Despite ambiguity in the NCSS position over whether schools should inculcate core values, the general tone of the NCSS paper is against the expectation of conformity in values, attitudes and beliefs concerning social issues.

The following is typical of the way values are treated in NAEP social studies objectives:

> V. Have a reasoned commitment to the values that sustain a free society.
>
> B. Believe in the freedoms of the First Amendment and can justify their belief (Social Studies Objectives, page 26).

NAEP social studies objectives are obviously consistent with the NCSS position that schools should deal with values. Furthermore, the term "reasoned commitment" implies something more than belief based on blind indoctrination. There are two dissimilarities, however, between the NCSS position and NAEP social studies objectives: (1) The social studies assessment is not ambiguous on whether certain core values should be inculcated; i.e., NAEP clearly expects children and young adults to be committed to certain values. (2) NCSS implies that a reasoned rejection of core values by students ought to be acceptable in American schools. There is no such implication in the social studies objectives of NAEP.

NAEP citizenship objectives. As with the social studies assessment, NAEP citizenship objectives agree with NCSS guidelines on the importance of dealing with values in school. But there is an even stronger tone tending toward values inculation in the citizenship assessment than in social studies. Although one citizenship subobjective states that rights are not absolutes and that they frequently conflict with each other (Citizenship Objectives, page 15), citizenship objectives which deal with values present them as

unqualified standards of proper behavior. Terms such as "reasoned commitment" and "justify their belief"are not appended to values objectives in the citizenship assessment as they are in social studies.

As previously indicated, NCSS stresses the importance of coming to grips with social problems, but does not advocate specific public policies or other solutions to social issues. In contrast, NAEP citizenship objectives sometimes advocate solutions, and they frequently state what NAEP considers to be morally correct stances. For instance:

> I.F. Understand and oppose unequal opportunity in the areas of education, housing, employment, and recreation.
>
> Age 17 . . .They are aware of the extent of unequal opportunity in their own communities, and they support legislation or other organized action to correct it, such as the "Headstart" program for underprivileged children (ibid., pages 13-14).

Numerous examples of the prescriptive tone of NAEP citizenship objectives could be given. One more should suffice:

> I.e. Are loyal to country, to friends, and to other groups whose values they share.
>
> Age 17 . . .They accept military service as an obligation of young men to help defend the nation's security. They feel allegiance to their country, as expressed in the Pledge Allegiance, respect for the flag, and other symbols (ibid., page 13).

Compared to NCSS guidelines, NAEP citizenship objectives overstress "correct" behavior, such as openly protesting to teachers about unjust school rules (ibid., page 19), but give too little attention to assessing the rational bases for the stance taken by a student. Social studies educators ought to compare the prescriptive tone of NAEP citizenship objectives with the following statement by NCSS:

> The ultimate power of rational processes resides in the explicit recognition of each person's opportunity to decide for himself in accordance with the evidence available, the values he chooses, and the rules of logic. Therein lies the link between human dignity and rational processes (Social Education, page 856).

Social studies educators should not forget that NAEP citizenship objectives attempt to assess rational processes, as indicated in a previous section of this paper. Nevertheless, in the majority of objectives which focus on values or on substantive issues, the emphasis is on whether the student takes the "correct" stance, rather than on whether he makes a rational choice. This emphasis on predetermined correct stance is out of harmony with the spirit of the NCSS guidelines.

Social Participation

NCSS guidelines and NAEP social studies objectives. Reference to social participation is limited to one subobjective and its age-level examples in the NAEP social studies assessment. That subobjective is:

> V.F. Are willing to participate in decision making relevant to their lives.
>
> Ages 9, 13 Participate in making decisions in school, at home, and in play groups.

Age 17 (1) Believe that, as much as possible, people should become involved in making decisions that affect them.

(2) Vote in school elections.

(3) Participate in making decisions in the groups to which they belong.

Adult (1) Believe that, as much as possible, people should become involved in making decisions that affect them.

(2) Vote in elections.

(3) Participate in political activity.

(4) Serve on juries when called (Social Studies Objectives, pages 27).

Apparently nothing in the above objective conflicts with NCSS guidelines on social participation. Nevertheless, striking omissions give that objective a milque-toast flavor compared to the NCSS posture. For instance, the most specific reference in the social studies assessment to participation in school-related decisions is that students should vote in school elections. Given the functions of most student governments, that hardly seems to be a critical example of civic involvement. In contrast, NCSS advocates that students participate in curricular decisions, stating bluntly:

> 1.1. Students should be involved in the formulation of goals, the selection of activities and instructional strategies, and the assessment of curricular outcomes (Social Education, page 860).

Furthermore, the central thrust of NCSS comments on social participation is absent from NAEP social studies objectives. That thrust includes: "Extensive involvement by students of all ages in the activities of their community. . . (ibid., page 859)." And, according to NCSS, social participation should be ". . .directed

toward the resolution of problems confronting society (loc. cit.)"

Limiting social participation to voting in school elections and participation in the affairs of peer groups falls short of meeting the intention of NCSS guidelines.

NAEP citizenship objectives. Compared to NAEP social studies objectives, far more attention is given to social participation in the citizenship assessment. Only one social studies objective focused on social participation. It is difficult to find a citizenship objective that does not include involvement by students in social problems. The following objectives and subobjectives are representative. The full sense of the objectives, however, cannot be appreciated without reading age-level examples, which are too long to reproduce.

I. Show concern for the welfare and dignity of others.

F. Understand and oppose unequal opportunity in the areas of education, employment, and recreation.

G. Seek to improve the welfare of groups of people less fortunate than they.

II. Support rights and freedom of all individuals.

. . .

III. Help maintain law and order.

. . .

D. Help authorities in specific cases.

. . .

V. Seek community improvement through active, democratic participation.

A. Believe that each person's civic behavior is important, and convey this belief to others.

B. Recognize important civic problems and favor trying to solve them.

C. Actively work for community improvement.

D. Participate in local, state, and national governmental process.*

. . .

VI. Understand problems of international relations.

. . .

B. Seek world peace and freedom for all peoples (Citizenship Objectives, page 9-32).

Acknowledgement that NAEP citizenship objectives come closer to NCSS recommendations on social participation than do NAEP social studies objectives ought to be tempered with the recollection that they are also strongly prescriptive; they are based on the assumption that correct social participation can be identified and ought to be expected. The following age-level examples for Objective III.E give that dual sense of active involvement and propriety:

Age 13 If they think a school rule is unjust they protest openly to teachers rather than covertly disobeying or grumbling.

Age 17 . . .Good 17-year-old citizens therefore use proper channels of appeal for proposing changes, e.g., class meeting, student council, teacher, principal, parents, PTA, school board, superintendent of schools, the courts, or elected representatives. They do not attempt unlawful means of change such as threats to the opposition, mob action, irrelevant slander or smear. Only if all lawful channels have been exhausted, and the injustice is quite serious, do they consider resorting to civil disobediance to try to change an unjust law or policy. . .(ibid., page 19-20).

---

*NAEP includes an age-level example of this subobjective only for adults, implying that it is inappropriate for students under eighteen years of age. That implication is contrary to the tone of NCSS guidelines. Children can perform most of the activities listed in the age-level example for adults.

One may wonder why responsible educators would object to prescriptive objectives such as the one above. Given the urban unrest of the 1960's--which should be remembered when explanations are given for the heavy emphasis on law and order in NAEP citizenship objectives--the above objective seems reasonable, even desirable. But upon close examination that objective is not as clear-cut as it first appears. Who is to decide when all lawful means have been exhausted? Are they to be exhausted separately for each issue that is raised? After long experience that lawful channels are unresponsive on one issue, is it reasonable to conclude that lawful protest on other issues would likewise be futile? Must lawful means be exhausted for each setting as well as each issue? For instance, was Martin Luther King, Jr., a poor citizen for not exhausting lawful protest in every city in which he led demonstrations? Is it reasonable to expect students to take serious complaints to Mickey Mouse student governments? To school administrators who never listen to students? To school boards who have repeatedly expressed support for authoritarian dress codes? In many instances, mild civil disobedience may be the most rational and responsible initial form of protest. At the very least, it is unrealistic to hold that good citizens must reach consensus on the above points.

The above discussion should not be taken as a brief for unlawful protest, but as a brief for the importance of assessing the reasoning behind the positions taken by students, rather than simply assessing whether those positions measure up to respectable expectations.

Before summarizing, one related point needs to be touched on. NCSS condemns the tendency for social studies to neglect viewpoints that represent either non-white or non-middle class groups (Social Education, page 856). Both NAEP assessments exclude middle class views of life at the expense of assessing other lifestyles and standards of behavior.* Educators who wish to pursue this point might begin by examining Objective VIII of the citizenship assessment. It simply restates elements of the protestant ethic, short of belief in God.

Summary

How good is the fit between the four areas of the NCSS rationale and the NAEP objectives for citizenship and social studies?

Knowledge. NCSS guidelines and NAEP social studies objectives differ on the manner in which social studies should be drawn from history and the social sciences. NCSS appears to define social studies primarily as citizenship education. NAEP social studies objectives appear to define social studies primarily as condensed replicas of history and the social sciences. Both agree that the academic disciplines are essential to social studies and that sources of knowledge other than history and the social sciences are important.

NAEP citizenship objectives are generally consistent with the NCSS position on the relation of academic disciplines to social studies. NAEP defines citizenship to include objectives that are neither referred to directly nor implied in NCSS guidelines.

*Emphasis on middle-class standards is not surprising, given the composition of NAEP review boards.

Abilities. NCSS stresses divergent thinking, data processing and human relations competencies. NAEP citizenship objectives attend to data processing skills, but are not as strong as social studies objectives in that area. Citizenship objectives are stronger than social studies objectives in human relations skills. Neither assessment emphasizes divergent-creative thinking. Furthermore, despite statements about the importance of dissent, the general tone of the citizenship assessment presses for conformity in values and in the stances taken on political-ethical issues.

Values. NCSS guidelines and both NAEP assessments agree that values should be dealt with in the school curriculum. Despite some ambiguity, NCSS opposes indoctrination of even basic values such as those contained in the Bill of Rights. NAEP social studies objectives on values include the phrase "reasoned commitment" and, therefore, appear to be closer to the NCSS position than are NAEP citizenship objectives. The latter generally prescribes correct belief and behavior.

Social participation. The NAEP social studies assessment contains one objective devoted to social participation. That objective, however, has a milque-toast flavor compared to NCSS guidelines and NAEP citizenship objectives. It is difficult to find an objective in the citizenship assessment that does not have some reference, direct or implied, to social participation. Examples of participation given in the citizenship objectives, however, are distinctly middle class to the exclusion of other viewpoints.

The following section deals with the last of three perspectives for answering the question: How good are NAEP social studies and citizenship objectives? The focus of that perspective is the question: How much overlap is there between or among sets of objectives?

Overlap Among Sets of NAEP Objectives

Preface

Objectives for at least five of ten NAEP assessments, including citizenship and social studies, contain enough overlap to justify comparison. The amount of overlap relevant to this examination varies from one or two subobjectives in some assessments to major portions of others.

Overlap is considered under five headings in the following discussion: democratic values, analytic competencies, human relations, knowledge and the protestant work ethic.

Democratic Values

With one minor exception, assessment of democratic values is limited to NAEP social studies objectives and citizenship objectives. Despite differences in wording, and despite important differences in the press for conformity to values, NAEP social studies and citizenship objectives offer nearly duplicate lists of democratic values. Both assessments stress substantive and procedural values, such as: freedom, dignity and worth of the individual; rights and freedoms contained in the Constitution, particularly the First Amendment; rule of law and due process. Both give special attention to equality of opportunity (Social Studies Objectives, pages 25-26; Citizenship Objectives, pages 9-16). Equality of opportunity is also the focus of Objective IV.D. (4) in the careers assessment (Career and Occupational Development Objectives, page 59).

Analytic Competencies

Five of ten NAEP assessments--Social Studies, Citizenship, Reading, Science, and Career and Occupational Development--give

substantial attention to analytic skills. In two of those assessments--Reading and Science--a majority of objectives deal with analytic competence. Although skills of analysis are not construed identically in each assessment, there is considerable genuine overlap. The following partial list of competencies mentioned in various assessment provides insight into the extent of overlap:

1. Identify and define problems

   (Social Studies Objectives, page 10; Career and Occupational Development Objectives, page 42; Science Objectives, page 13)

2. Recognize or develop hypotheses

   (Social Studies Objectives, page 10; Science Objectives, page 14)

3. Obtain information relevant to the problem

   (Social Studies Objectives, page 10; Career and Occupational Development Objectives, page 43; Science Objectives, page 22)

4. Detect logical errors

   (Social Studies Objectives, page 11; Citizenship Objectives, page 35; Career and Occupational Development Objectives, page 42; Reading Objectives, page 23)

5. Recognize emotive loading, rhetorical devices and propaganda techniques

   (Citizenship Objectives, page 35; Career and Occupational Development Objectives, page 42; Reading Objectives, page 14, 17, and 23)

6. Distinguish facts from opinions, relevant from irrelevant information, and reliable from unreliable sources

   (Social Studies Objectives, page 11; Reading Objectives, page 23; Science Objectives, page 18)

7. Detect unwarranted assertions, generalizations, conclusions and unstated assumptions

   (Social Studies Objectives, page 11; Career and Occupational Development Objectives, pages 33-34)

Some assessments give special emphasis to different aspects of analytic thinking. The science and reading assessments, for instance, have notably different content in those objectives dealing with thinking skills. One unique aspect of the science assessment is a set of subobjectives which describe the relation between observation, laws, and theories (Science Objectives, pages 19-21). Science also has the most complete set of objectives about the generation and validation of hypotheses (ibid., pages 13-19) and is the only area to explicitly recognize the importance of suspended judgement.

The detail with which some traditional sources of critical thinking concepts are dealt with is a distinctive characteristic of the reading assessment. They include inductive and deductive reasoning (Reading Objectives, page 21), rhetorical techniques (ibid., page 23), logical falacies (ibid., pages 23-24), propaganda devices (ibid., page 23), connotative and denotative meaning (ibid., page 14), and emotive appeals (ibid., pages 17-19). No other assessment devotes explicit attention to the critical analysis of written materials, even though books and articles are important sources of information in most school subjects.

Surprisingly, one of the most detailed treatments of analytic competencies is found in career and occupational development objectives. It may be the only assessment to specify skill in interpreting statistical data (Career and Occupational Development Objectives, page 33).

Human Relations

Overlap occurs in three of ten assessments: Social Studies, Citizenship, and Careers and Occupational Development. Social studies

objectives imply the importance of human relations skills, but do not specify which ones are important. Aside from noting that students should be sensitive to the thought and feelings of others (Social Studies Objectives, page 10) and should participate in decision making in groups to which they belong (ibid., page 27), competence in interpersonal relations is neglected.

Compared to the social studies assessment, citizenship objectives give more detailed indications of which human relations skills are important. One subobjective is especially relevant: V.E. Apply democratic procedures on a practical level when working in a group (Citizenship Objectives, page 29-30). But the most detailed listing of interpersonal skills is given in the careers assessment (Career and Occupational Development Objectives, pages 46-68). Among the items stressed are: giving and receiving criticism; obeying reasonable authority and expressing minority viewpoints; sensitivity to nonverbal communication such as expressions, tone, and gestures; considering consequences for others of group decisions; working toward common goals.

Knowledge

Viewed as information drawn from traditional academic disciplines, there is very little overlap in knowledge among the contents of the ten assessments. The exception which is most relevant to this paper was noted in a previous section. Both the social studies and citizenship assessments contain objectives which are drawn from political science and focus on the structure and functions of government (Social Studies Objectives, pages 24-25; Citizenship Objectives, pages 21-25). Of the two, the citizenship objectives are more detailed.

The Protestant Work Ethic

The relevance of this topic--the protestant work ethic--may not be apparent to social studies educators. It is included because an understanding of overlap between the citizenship and careers assessments is essential to understanding recommendations which will be made later in this paper.

Previous sections of this paper noted that some NAEP citizenship objectives focus on content not mentioned in NCSS guidelines, would be considered by many teachers to be outside the domain of social studies, and might be viewed as only tangential by others. It was also noted that citizenship objectives reflect a middle-class frame of reference to the exclusion of other frames and tend to prescribe correct behavior rather than rational decision making. Consider the following objective and subobjectives:

> VII. Take responsibility for own personal development and obligations.
>
> A. Further their own self-improvement and education.
>
> B. Plan ahead for major life changes.
>
> C. Are conscientious, dependable, self-disciplined, and value excellence and initiative.
>
> D. Economically support self and dependents (Citizenship Objectives, page 39-43).

The total set of age-level examples for the above subobjectives is too long to reproduce. The tone of those examples is evident, however, in the following:

> Age 17 They exploit and develop their talents to the maximum, and they seek to learn as much as they can rather than to get by with the least effort. They explore different fields to better learn their own interests and talents. They schedule time spent on studies and on different types of recreation. . . .

> . . .They take advantage of our free enterprise system by inventing, producing, or marketing useful products and services. They seek part-time or summer jobs. . . . They start new tasks without having to be told, check work for mistakes, and use initiative to find better ways of achieving work goals. They are not careless or tardy in keeping appointments. . . (Citizenship Objectives, pages 40 and 42).

Much of the content and some of the tone of the above objectives are duplicated in the careers assessment. Although the tone is oriented toward middle-class work values, there is more emphasis on choosing among alternatives. A partial list of relevant objectives and subobjectives includes:

I. Prepare for making career decisions.

A. Know own characteristics relevant to career decisions.

1. Be aware of own current abilities and limitations.

2. Be aware of own current interests and values.

. . .

C. Relate own personal characteristics to occupational requirements.

D. Plan for career development or change.

. . .

II. Improve career and occupational capabilities.

Age 17 (1) Pursue education and training. . . .

(2) Obtain part-time and summer work. . . .

. . .

(5) Study on own initiative. . . .

. . .

IV. Practice effective work habits.

A. Assume responsibility for own behavior.

. . .

B. Plan work.

. . .

C. Use initiative and ingenuity to fulfill responsibilities.

. . .

Age 13 Do more than required or expected, such as reading beyond assigned pages, asking to make up work that they have missed, and volunteering to learn special skills.

. . .

Think of ways to perform part-time work more quickly and more effectively. . . .

. . .

V. Have positive attitudes toward work.

. . .

B. Hold competence and excellence in high regard.

. . .

D. Value work in terms of societal goals.

. . .

Age 17 (1) Believe that each person should strive to be self supporting to the extent that he is able (Career and Occupational Development, pages 16-58).

There is also some overlap between the careers and citizenship assessments concerning health and safety practices (Career and Occupational Development Objectives, pages 54-55; Citizenship Objectives, page 11). The careers assessment construes them as part of good work habits. Citizenship objectives construe them as showing concern for the welfare of others.

Summary

The purpose of this section was to indicate the extent of overlap on topics relevant to citizenship and/or social studies. Five such topics were identified and overlap was indicated among five of ten NAEP assessments. The amount of overlap varies in a number of ways, including: the number of assessment areas

involved; the amount of duplication among assessments; and the amount of attention given to an overlapped topic by different assessments.

Most of the overlap on democratic values is limited to social studies and citizenship objectives, which give different emphases to virtually duplicate sets of values. The most pervasive area of overlap concerns analytic competencies; half of all NAEP projects attend to this area of the curriculum. Three assessments deal with human relations skills. Listed in ascending order concerning the amount of attention devoted to interpersonal relations, they are: social studies, citizenship, and career and occupational development. The least amount of overlap among assessments involves traditional content drawn from the academic disciplines and is limited to citizenship and social studies objectives. The protestant ethic is given substantial attention in the citizenship and careers assessments.

Three assessments, other than citizenship and social studies, should be of interest to social studies teachers because each assessment approaches skills of analysis from different perspectives. One focus of the reading assessment is the analysis of written claims. It also draws more heavily from logic than do other assessments, and includes concepts from traditional critical thinking categories such as propaganda analysis and rhetoric. The science assessment is the only one to deal with that view of scientific process which stresses the interplay of observation and theory. Social studies educators who believe that students

should be taught powerful ideas from the social sciences may want to examine that part of the science objectives. The careers assessment should be of interest to social studies educators because of the amount of detail devoted to analytic competencies which range from defining problems to devising and implementing alternative courses of action.

Whether overlap among assessments is reasonable will be dealt with in the following section on recommendations.

Recommendations

The following suggestions are intended for two audiences. The first is social studies educators who wish to examine results of NAEP assessments for information about how well American students meet various objectives. The second audience is NAEP staff and contracting agencies.

Social Studies Educators

Educators who note that the social studies and citizenship assessments do not completely satisfy NAEP's criteria for objectives, are not completely consistent with NCSS curriculum guidelines, and overlap at least three other assessments may conclude that social studies specialists have little to gain by examining NAEP objectives and exercises. That conclusion would be an error for a number of reasons.

First, NAEP criteria and NCSS guidelines are not consistent with each other. NAEP criteria demand that objectives reflect actual school practice. NCSS guidelines look to how practice can be improved. To the extent that NAEP objectives are consistent with one of those standards, they must violate the other. That inconsistency is not necessarily a weakness in NAEP, but is partly a function of examining NAEP objectives from two different perspectives.

Second, it is doubtful that a cross section of teachers would each give equal weight to the three perspectives from which we have viewed NAEP objectives.

Third, it is likely that nearly any social studies teacher will find several NAEP objectives which tap something which that

teacher believes is worthwhile.

The first recommendation for teachers, therefore, is that they pick and choose among NAEP citizenship and social studies objectives; that they neither accept or reject objectives without reflecting on how each measures up to the teacher's view of what is worth teaching. The second recommendation is that social studies educators not limit their examination to the social studies and citizenship assessments. Many objectives in reading, science, and career and occupational development are relevant to social studies education. Even the art assessment has one objective which may interest social studies teachers (Objective IV.E, Ages 17, A in Art Objectives, page 18).

NAEP Staff and Contractors

NAEP citizenship and social studies objectives have been critiqued from three perspectives in previous sections of this paper:

1. Consistency between NAEP objectives and NAEP criteria;
2. Consistency between NAEP objectives and the NCSS position paper on curriculum guidelines;
3. Overlap in objectives relevant to citizenship and social studies among five assessments.

What implications are indicated by each of those perspectives? Inconsistency between NAEP objectives and NAEP criteria can be handled either by changing the criteria, or by changing the objectives, or both. Later, recommendations will be made for modifying objectives. At this point it is recommended that NAEP seriously consider changing at least one criterion: the requirement that objectives reflect goals which schools are seriously trying to achieve. A national assessment of social studies can serve

professionals, laymen and students by measuring achievement both in that which is generally taught and that which should be taught but generally is not. Of course, it is much easier to obtain agreement on what generally is included in the curriculum than on what ought to be.

Modification of NAEP criteria would also be one way to handle inconsistency between NAEP objectives and the NCSS position paper. At the present time, objectives cannot be consistent both with the NAEP criterion which requires that objectives reflect actual school practice and those NCSS guidelines which look to the improvement of practice. NAEP criteria should allow a reasonable mixture of both types of objectives.

The above suggestion may also help NAEP handle a sticky problem, which is how to anticipate and stay abreast of school practice. To the extent that social studies curricula are not stagnant, NAEP objectives will always be behind times if assessment is restricted to that which is taught in schools at the time objectives and exercises are developed. Including objectives from the vanguard of

social studies may help close the gap between assessment and current practice, assuming that elements of the vanguard find their way into instruction within the time covered by an assessment cycle.

Two specific examples of possible application of the above recommendation can be given. The first involves Objective IV of the social studies assessment: Have knowledge relevant to the major ideas and concerns of social scientists (Social Studies Objectives, page 13). This objective reflects current practice regarding selection of content from history and the social sciences. It could be modified to also reflect the NCSS view that content should be selected for relevance to social issues and the concerns of students.

The second example involves the prescriptive tone of the majority of citizenship objectives. Although schools tend to ·escribe values, rather than help students clarify thier own ·nitments, in this case prescription is not consistent with ·rrent school practice; it is doubtful that schools seriously attempt to deal with basic democratic values, even at the level of indoctrination. Nor is prescription consistent with NCSS guidelines. It is recommended that greater emphasis be given to the assessment of rational analysis and that less emphasis be given to correct stances on issues and values. This recommendation is consistent with current thinking on the improvement of practice as that thinking is reflected in the NCSS position paper.

The third perspective--overlap among assessments--provided implications for revising objectives, but not all overlap is

excessive. For instance, it is reasonable for one art objective to focus on making inferences about social structure from art produced by a society. Nor is the careers assessment guilty of unnecessary duplication when it includes an objective on equal opportunity in employment. In both cases, the amount of overlap with objectives in citizenship and social studies is minor.

Some overlap, however, is wasteful duplication. It does not seem reasonable for social studies and citizenship objectives to duplicate assessment of democratic values. Nor does it seem reasonable for both to include substantial sections which assess knowledge of the structure and function of government. It has also been noted that they include overlap in the assessment of analytic skills.

Overlap could be eliminated by restricting objectives on a given topic to a single assessment. If social studies dealt with analytic skills, for instance, citizenship would not. If citizenship dealt with democratic values, social studies would not.

That alternative is not recommended, however, for the following reasons. First, a citizenship assessment which did not include political-ethical values would not make sense. It could not assess competence in a major element of democratic citizenship. Second, a social studies assessment which ignored political-ethical values would likewise make little sense to many social studies educators. Social studies in the United States is generally held to have a special obligation for political citizenship. Recognition of that obligation, for instance, is reflected in the request that NCSS critique both the NAEP social studies and citizenship projects.

It is, therefore, recommended that NAEP consider major revision of the social studies assessment and elimination of the citizenship project. This revision would require incorporating into social studies those portions of the citizenship assessment which are most relevant to political citizenship. Little of the present citizenship assessment would be lost by this move because many of the citizenship objectives which do not assess political citizenship are contained in objectives of the career and occupational development project. In other words, most of the important elements of the citizenship assessment--political-ethical values, analytic skills, human relations skills, democratic participation, work values, and health and safety practices--are either now contained in the social studies or careers assessments or would be incorporated into social studies.

References

National Assessment of Educational Progress. Art Objectives. Ann Arbor, Michigan: NAEP, 1971.

_____. Career and Occupational Development Objectives. Denver: NAEP, 1971.

_____. Citizenship Objectives. Ann Arbor, Michigan: NAEP, 1969.

_____. The National Assessment Approach to Exercise Development. Ann Arbor, Michigan: NAEP, 1970.

_____. Reading Objectives. Ann Arbor, Michigan: NAEP, 1970.

_____. Science Objectives. Ann Arbor, Michigan: NAEP, 1969.

_____. Conferences on the Development of Instrumentation for Assessment of the Progress of Education in U.S. Minneapolis, Minn., NAEP, 1967. (Mimeo Report)

_____. Social Studies Objectives. Ann Arbor, Michigan: NAEP, 1970.

_____. What is National Assessment? Ann Arbor, Michigan: NAEP, 1970.

NCSS Task Force on Curriculum Guidelines. Social Studies Curriculum Guidelines, Social Education, 1971, 35, 853-869.

Newman, Fred M. and Donald W. Oliver. Clarifying Public Controversy. Boston: Little, Brown and Company, 1970.

Oliver, Donald W. and James P. Shaver. Teaching Public Issues in the High School. Boston: Houghton-Mifflin Company, 1966.

CRITIQUE OF NAEP PROCEDURES

TASK I

FINAL REPORT PART TWO

A. Guy Larkins

University of Georgia

November, 1974

SE 008 020

This report represents part of an investigation funded by the Education Commission of the States, and coordinated by a special Steering Committee working under the auspices of the National Council for the Social Studies.

Introduction

The purpose of this critique is to help social studies educators determine the extent to which the findings of National Assessment of Educational Progress (NAEP) are grounded in sound practice. The final report for Task I was written in two parts. The objectives upon which NAEP based its test items were discussed in the first paper, Part One of the final report. The other criteria and procedures used to develop citizenship and social studies exercises are discussed in the first section of this paper, Part Two.

The following pages deal with several topics: instrumentation, sampling, data analysis and procedures used to report NAEP results. For several reasons, a more detailed discussion is presented of the first topic, instrumentation, than of the others. The development of valid measures was perhaps the most crucial problem facing NAEP. Exercise validity also seems to be an area in which human judgment, with all of its tendency toward disagreement, plays a pre-eminent role. Other problems, such as sampling and data analysis, also call for informed judgment, but seem to be more amenable to solutions upon which experts can agree.

Instrumentation

This portion of the critique summarizes and comments on the procedures used by NAEP to develop assessment instruments. The central focus of this section is whether NAEP used adequate procedures. The focus is not on whether NAEP exercises are valid, but on possible strengths and weaknesses of the procedures used to develop those exercises. The validity of NAEP exercises is dealt with elsewhere (Hunkins, 1973).

Exercises compared to tests

The term "exercise" appears frequently during the following discussion and requires some explanation. NAEP did not develop the usual type of achievement test. An achievement test is generally thought of as a set of items, correct responses to which can be summed to yield scores. Such tests are usually used to make judgments about the performance of individual students. In contrast, NAEP developed single-item instruments which they call "exercises." Instead of reporting scores for individuals, or average scores for groups, NAEP intended to report percentages of correct response to individual exercises. The following is an example paraphrased from an NAEP report: Forty-five percent of the sample of nine-year-old children gave acceptable reasons for why a newspaper

BEST COPY AVAILABLE

report can be wrong (Citizenship, National Results, 1970, page 103).

Because each NAEP exercise is a single-item test, the validity of an exercise is probably more crucial than the validity of any one item in a multiple-item achievement test. In multiple-item instruments, there is some room for slippage. Several good items may partly compensate for an occasional bad one. But when single-item exercises are used, each invalid exercise leads to an invalid finding.

Reviews relevant to validity

NAEP exercises were based on objectives developed for each of the ten assessment areas. Following the development of objectives, the production of exercises was turned over to private contracting agencies. When the contractor delivered exercises to NAEP, a series of reviews was conducted. Some reviews focused primarily on problems such as clarity, meaningfulness, and offensiveness of exercises. But some dealt explicitly with content validity, and several of those which did not focus directly on validity did so indirectly.

Originally, NAEP planned two reviews. In the first, laymen inspected those exercises which the NAEP staff thought might be offensive to the public, even if the probability of

offensiveness seemed slight. Exercises which passed the offensiveness review were given a second review by subject matter specialists. At first, NAEP thought that these two reviews would suffice, but found that several additional review conferences were required (Finley & Berdie, 1970, p. 35). Those various reviews are discussed below.

The lay reviews. One of the innovative features of NAEP is that laymen were asked to judge the quality of assessment instruments. The involvement of laymen posed some special problems. The first concerned the selection of participants. When educators are asked to review achievement instruments they are chosen on the basis of some presumed technical competence. In contrast, the laymen were chosen not for technical competence, but to help insure that the exercises would be well received by parents and other concerned citizens. NAEP attempted to obtain a sample of such concerned citizens by seeking nominations from such organizations as the AFL-CIO, the U.S. Chamber of Commerce, the NAACP and others (Ibid., p. 38).

As a result of the selection procedure, the lay review panels probably contained a high proportion of middle-class, college educated persons who were active in civic affairs. An argument in favor of such panelists is that they are probably

similar to those persons who control access to schools; eleven of the seventeen organizations which nominated panelists have some varient of the word "education" in their titles. Access to schools, of course, is essential to an organization like NAEP.

On the other hand, if the primary reason for using lay panels is to obtain feedback from a broad spectrum of parents and concerned citizens, other selection procedures should have been used. Personal communication between the author and NAEP staff members indicates that they are aware of the problem and have taken steps to broaden the representation on their lay committees.

A second problem was that laymen have not generally played a formal role in evaluating assessment instruments. There were, therefore, no clear traditions established concerning the function which they might serve. NAEP assigned them the task of reviewing exercises for possible offensiveness.

Although screening exercises for offensiveness is a useful role for laymen, it may be too narrow a definition of their task. For instance, one of the nine recommendations of the first lay conference was directly concerned with content validity. That suggestion was:

> Some areas are more complex than the exercises suggest. Possibly open-ended questions would be more effective, (examples: "What is the main purpose of the United Nations?" and "Why is it desirable to have two newspapers available in a town?") or open-ended interviews (Summary of Discussion, Conference for Lay Review of Sample Exercises, Chicago, May 10-12, 1966).

By virtue of their education or experience, panelists probably can make useful suggestions about content validity and the appropriateness of exercises for students of various ages, ethnic backgrounds or social classes. This suggestion does not mean that lay judgments should replace those of professional educators, subject matter specialists or experts in measurement.

Although the lay panels were not asked to review exercises for content validity, their judgments probably affected the validity of exercises in at least two ways; the first is the possibility that manipulating the offensiveness of an item might either increase or decrease its validity. For instance, one aspect of content validity is whether an exercise adequately represents some information, skill or attitude which is considered part of the content to be tested. It might not be possible to test some parts of a legitimate social studies or citizenship content without getting into sensitive topics. This is particularly true when assessing attitudes or values. In those cases, the lay review might cause exercises to be so

modified that a decrease in offensiveness is accompanied by a decrease in validity.

A problem which is closely related to the one above is whether a total body of content has been adequately sampled. The removal of exercises deemed offensive may result in less than adequate sampling of the total domain of citizenship or social studies.

The potential danger to validity is illustrated by the following selection of some of the topics which the lay panels found offensive. The topics listed are especially relevant to citizenship and social studies:

> References to specific minority groups should be eliminated whenever possible.
>
> * * *
>
> The area of religion was considered so sensitive that it could not be handled without emotional reaction.
>
> * * *
>
> The overabundance of exercises dealing with a person's rights was considered offensive unless more exercises were added dealing with his responsibilities in a free society.
>
> Any reference to sex, unwed mothers, divorce, whiskey, the FBI, the President, Communism and specific organizations such as the Klu Klux Klan and labor unions, might make an exercise offensive unless extreme care was used in the wording.

* * *

> Exercises which show national heroes in an uncomplimentary fashion though factually accurate are offensive.
>
> Exercises which might be interpreted as putting the police or other authorities in an unfavorable light are offensive (Finley & Berdie, 1970, pp. 42-47).

Examples could also have been selected which might improve validity. For instance, some social studies educators do not consider private issues to be a legitimate concern of the school curriculum. Nor would increased ethnocentrism be considered a proper social studies objective. The following concerns voiced by the lay panels may, therefore, help improve the validity of citizenship and social studies exercises:

1. Invasion of privacy	Questions in any way connected with family financing. Included would be questions dealing with how much an individual earned or received as an allowance, how income is budgeted, what contributions are made and what taxes are paid.

* * *

2. Inferiority of other nations	Any exercise implying the inferiority of other nations or exercises which imply the superiority of Americans to people in less well developed countries are offensive (Finley & Berdie, 1970, pp. 42 and 46).

NAEP recognized the potential conflict between validity and offensiveness and asked the lay panels to weigh the importance of questionable exercises against the possibility that the public would object to them (Ibid., p. 36). Theoretically, at least, an extremely important exercise might be included in the assessment even though it might seriously offend someone. In practice, it appears that most of the potentially offensive exercises were revised to exclude the offensive wording.

It is impossible to judge the extent to which the lay reviews affected content validity. But that the potential impact was considerable can be seen from the following data: Of 1776 social studies exercises, 185 were reviewed for offensiveness, 79 were dropped as a result of the review, and 28 were revised. Of 600 citizenship exercises, 305 were reviewed by laymen, and only eight were dropped as a result. But, "Citizenship exercises were reviewed and revised and re-reviewed so often that no figure [on the number revised as a result of the lay conference] is meaningful . . . (Ibid., p. 40)."

The following information throws additional light on the above data. Of five lay conferences, the first two dealt exclusively with citizenship exercises, the third dealt with citizenship and social studies in addition to literature and

science, the fourth examined exercises in citizenship, social studies, reading, art and literature, and the fifth considered citizenship, social studies, vocational education, art, literature, music and reading (Conferences on the Development of Instrumentation . . . June 15, 1965-June 30, 1967). In other words, every lay conference examined citizenship exercises, and three of the five examined social studies. Only literature came close to receiving as much attention as the two areas most relevant to social studies educators. Our profession, therefore, should be particularly interested in the possible impact of the lay panels on content validity.

It is the judgment of this author that NAEP could have reduced the potential danger to content validity by having social studies specialists review both the items which passed the lay review and those which did not. In some cases, subject matter specialists might have been in a better position than the lay panels to weigh the importance of an exercise against its potential offensiveness. Although the subject matter specialists later approved of most of the exercises which passed the lay review, they may not have approved of dropping or revising some of the exercises which were rejected by the laymen.

Despite the reservations voiced above, it is the judgment

of this author that the lay reviews were a helpful innovation in exercise development. The most obvious payoff came when the exercises were administered. Apparently, very few schools which were asked to participate in the first round of assessments found the exercises objectionable (Finley & Berdie, 1970, p. 47).

Additional checks on offensiveness and importance. A mail review by subject matter specialists followed the first round of lay conferences. Before discussing the subject matter conferences, however, it may be useful to jump ahead and mention additional ways in which NAEP involved laymen and checked on the offensiveness and importance of exercises.

As previously pointed out, NAEP was concerned about the potential conflict between offensiveness and importance of assessment instruments. Their concern was two-edged: Some important items might be offensive and therefore unwisely excluded from the assessment. And some unimportant items might be inoffensive and therefore unwisely included. As a check against the second concern, laymen were asked to review exercises for meaningfullness at a late point in the development of exercises. Although the exercises had successfully passed a number of previous reviews, about sixteen percent were judged by the laymen to be trivial (Ibid., p. 101). Since this seemed

to be a substantial rejection rate, the Technical Advisory Committee of National Assessment also reviewed the exercises for offensiveness and importance, as well as for validity and clarity (Ibid., pp. 103-106).

A final review was conducted by the United States Office of Education. Supposedly, this review was limited to invasions of privacy. But the main concern may have been with the broader problem of offensiveness. The following question was dropped from the science assessment because of objections by USOE: How are the components of contraceptive pills which contain estrogen and progestine intended to function to prevent conception (Ibid., p. 107)? Although this question may offend some parents, it does not probe into a person's private life.

Social studies specialists should be interested in the fact that of four exercises which were dropped as a result of the USOE review, three were in citizenship. Furthermore, all of the eleven exercises which were modified as a result of this review were in citizenship. This does not mean that the USOE review was harmful. At least some of the citizenship exercises did infringe the privacy of students (Ibid., pp. 107-108).

Mail review by subject matter specialists. Following the lay reviews, exercises were sent to subject matter specialists

nominated by professional organizations such as the National Council for the Social Studies. Reviewers were asked to consider whether the exercises sampled the objective indicated, whether they contained flaws such as ambiguous wording, and whether the reviewers agreed with the answer indicated in the key. Estimates of the difficulty level of each exercise for a given age group were also requested.

As a result of this review, less than one percent of the exercises were rejected, but approximately twenty-five percent needed revision. An interesting sidelight is that only seventeen percent of the citizenship exercises received negative comments by the subject matter specialists, compared to thirty-two percent for social studies and fifty-two percent for reading (Ibid., p. 49). Perhaps the repeated scrutiny given the citizenship exercises by the lay panels resulted in a more polished product than was true of other subjects. If so, it suggests that the lay panelists improved content validity, even though their charge was limited to screening exercises for offensiveness.

Although the mailed review proved helpful in identifying problems with content validity, it contained several weaknesses which NAEP attempted to correct in later reviews: (1) The number of reviewers was small. Only two persons took part

in the citizenship review (Summary of Development of Citizenship--1965-1969), and each exercise was examined by only one reviewer. Greater diversity in professional judgment might have been helpful. (2) Each reviewer received only those objectives upon which the exercises he critiqued were based. It was impossible, therefore, for reviewers to determine whether the larger content of an assessed subject was adequately sampled. (3) Communication between the reviewers and the exercise developers was hampered by lack of face to face contact (Finley & Berdie, 1970, pp. 51-52).

It was originally intended that the exercises would go through only the two reviews outlined above. But at this stage, it was obvious that more extensive revision was needed than was originally anticipated. Therefore, NAEP decided to hold a series of subject matter review conferences.

Subject-matter review conferences. The involvement of social studies specialists in the validation of exercises was much more extensive than originally planned. In addition to the original mail review, at least four conferences were held to review exercises for the social studies or citizenship assessments. These conferences were attended by a small number of professionals with expertise relevant either to the subject matter, the age-level and background of students, or

test development. The first two conferences were held in June of 1970 and January of 1971. They focused on both social studies and citizenship (Item Development: Social Studies). Two other conferences, held in August and October of 1970, were limited to social studies (Social Studies Exercise Review . . . 1970, and National Assessment of Educational Progress Subject Matter Review, October 8-11, 1970). A conference was also held in February and March of 1968 to produce additional social studies exercises for thirteen-year-olds. And a similar conference was held in April of 1970 to produce additional exercises which would be relevant to young people from minority groups (Social Studies Writing Conference . . . 1970).

These conferences illustrate the determination of NAEP to improve the quality of exercises. When the mailed reviews were not adequate, NAEP held conferences which would allow reviewers to meet face to face. When the early conferences indicated a need for greater attention to assessing the achievement of minority group students, additional writing and review conferences were held. After each review, questionable exercises were returned to the contractors for modification. In some cases, new exercises were produced. Each time new exercises were developed, those items were also reviewed.

In short, it appears that NAEP went to considerable trouble

to involve subject matter specialists in reviewing exercises. In terms of sheer number of reviews, it is difficult to imagine how NAEP could have given greater attention to face validity.

However, NAEP may have been able to improve the subject-matter reviews by using more systematic procedures. Panelists at NAEP's subject-matter conferences appear to have been given short lists of criteria for judging exercises. In at least some conferences, panelists were also given global rating scales (Finley & Berdie, 1970, pp. 67, 68 and 72). These procedures, however, do not appear to have been as thorough or systematic as those used by Hunkins (1970, pp. 15-18).

That the reviewers on Dr. Hunkins' panel were given a more systematic and detailed procedure to follow when reviewing exercises may account for his finding that several NAEP objectives " . . . are not represented by any exercises or are represented by exercises that have been judged not valid (Ibid., p. 92)." There are, however, at least two other plausible explanations for Hunkins' judgment that some NAEP exercises are invalid: (1) Face validity is based on human judgment; well-qualified and well-intentioned people will disagree. (2) That some instances of invalidity remained in NAEP exercises may have been due to the difficulty in writing good items in some

areas, rather than in identifying which areas needed better items. Nevertheless, NAEP should investigate the possibility of using more detailed and systematic procedures for reviewing exercises.

Field studies

Several NAEP criteria for exercises have been mentioned in the above discussions. In addition to reviews which focused explicitly on content validity, exercises were reviewed for possible offensiveness, invasion of privacy and importance. At least three other criteria were given serious attention. They are clarity, difficulty level and directionality. Clarity and difficulty level are discussed in this section.

An exercise may fail to meet the criterion of clarity in several ways: Students may not be familiar with the format of a test, or instructions may be confusing. The wording of an exercise may be vague or ambiguous, or the vocabulary may be too difficult. Any of these problems may confound measurement of an extraneous variable with measurement of knowledge of the content of the exercise.

NAEP specified that each exercise must be written to meet one of three difficulty levels--very easy, moderately difficult, or very difficult. An easy item is one that approximately ninety percent of the students can answer. A difficult

one can be correctly answered by roughly ten percent (Finley and Berdie, 1970, 28-30).

Before discussing the field studies, two additional points need to be made: (1) Both clarity and difficulty of exercises were examined in the various reviews, but they also occupied a prominent role in the field studies. (2) Clarity and difficulty level are intertwined; lack of clarity can change an otherwise easy exercise into a difficult one.

The 90 percent study. This study attempted to determine whether those exercises which were written to be very easy could be answered by approximately 90 percent of the assessees. A sample of slightly more than 800 students was divided among two age levels, 9 and 17, two socio-economic classes, high and low, and two areas of the country, the northeast and southeast. Among the 646 exercises which were tried out, 84 were from social studies. Citizenship was the only assessment area not included in this study.

Although the sampling procedures in the 90 percent study do not allow for strict generalization to the population which NAEP wished to assess, the results provide convincing evidence that exercise writers missed the 90 percent criterion by a wide margin. The median difficulty for "easy social studies exercises was 67 percent at age 9, and 60 percent at age 17.

For both age groups, social studies exercises which were written to be very easy covered nearly the full range of possible difficulty levels (Finley & Berdie, 1970, pp. 55-60).

The 90 percent study apparently was relevant to the criterion of clarity. In attempting to account for the unexpected difficulty of "easy" items, NAEP focused on problems of communication and vocabulary (Ibid., p. 60). The subject-matter conferences, which have been discussed earlier in this paper, were in part an outgrowth of the concern for clarity which was reinforced by the 90 percent study.

Feasibility studies. Two feasibility studies were conducted. The in-school study sampled low-achieving students in grades three, seven and eleven. In the out-of-school study, 17-year-old youth and young adults were interviewed in their homes. Both studies provided important information about the clarity of exercises.

The in-school study included exercises which NAEP identified as being potentially difficult for students to understand. After the packages of exercises were administered, a few of the lowest achieving students in each class were interviewed to help spot difficulties in understanding test instructions, format, vocabulary and vague or ambiguous terms.

No hard data were reported, but several problem areas

were either identified or re-emphasized. Exercises apparently needed additional editing for clarity and simplicity. The number of exercise formats needed to be reduced. And procedures for reducing reading difficulties needed to be worked out. It was suggested, for instance, that exercises be read aloud (Ibid., pp. 60-66).

No information concerning sample size, geographic distribution, or sampling procedures is given for these studies. Finley and Berdie (1970, p. 60) simply refer to them as "the first large-scale tryouts."

Citizenship study. A small-scale field study was conducted to identify citizenship exercises which low achieving students would find difficult to understand. Eleven boys and girls, ages 9, 13 and 17 were given all of the citizenship exercises. Then the children were interviewed to identify exercises which were unclear. One of the results dramatizes the importance of determining whether students understand test questions: "At the 9-year-old level, the three students were such poor readers that each exercise had to be read aloud before they were able to answer (Ibid., p. 73)."

Mathematics and choices studies. Although the subject-matter for the first of these investigations was mathematics, the findings concerning clarity and difficulty of test

formats may also hold for social studies and citizenship. The subject-matter for the second study was social studies.

One of the objectives of both investigations was to determine how changes in format affect the difficulty of exercises. For instance, are open-ended* exercises more difficult than multiple-choice items? Can the difficulty of an item be changed by manipulating the distractors? Does the inclusion of "I don't know" as an option on multiple-choice questions reduce guessing?

The mathematics study also investigated the effect of substituting simple language for technical terms in an attempt to increase clarity.

Questions about the effect of different formats on the difficulty of exercises are especially important. If item difficulty is due to both format and content, it may be meaningless to compare results by individual exercises unless identical formats and equally difficult distractors are used. In other words, NAEP results may confound knowledge of content with an extraneous variable, difficulty of format.

---

*By open-end, NAEP apparently means a short answer exercise, regardless of whether it is intended to solicit a divergent or convergent response.

The mathematics study used six classes of high school-age students. The classes ranged in knowledge of math from very poor to advanced. Four packages of exercises were randomly assigned to students in each class. Each package contained 16 exercises with technical wording and 16 with simplified wording. The content of exercises in the four packages were identical, but each package was limited to one of the following formats: (1) Multiple-choice with an "I don't know" option, (2) Multiple-choice without "I don't know," (3) Open-end with an "I don't know" option, and (4) Open-end without "I don't know (Ibid., pp. 77-78)."

The choices study used 11 classes of junior high school students of average and below average ability. Four packages of exercises with identical content, but different formats, were randomly assigned to students in each class. Three of the formats were multiple-choice with distractors manipulated to produce different levels of difficulty. The fourth format was open-ended. Each format included an "I don't know" option (Ibid., pp. 81-82).

Random assignment of individual students to treatment in each study is one mark of good internal validity. No systematic differences should exist among students who received different packages. And, although the sampling procedure does

not allow for strict generalization to other populations, the results summarized below are both interesting and important.

The following results are combined from both studies: (1) Whether technical math terms or everyday terms were used made little difference. This finding was taken by NAEP as support for their decision to emphasize simplicity of language. (2) Format had little effect on the responses to very difficult or very easy exercises. But for items of moderate difficulty, fewer students gave correct responses to open-end questions than to multiple-choice items. And the difficulty of about two thirds of the multiple-choice questions was influenced by variations in the distractors. (3) Difficulty of multiple-choice exercises with the "I don't know" option was closer to that of open-end questions than to ordinary multiple-choice items (Ibid., pp. 79-80 and 83-85).

Readers of NAEP reports should keep the above results in mind when drawing conclusions such as "More nine-year-olds can do X than Y." Differences in performance on exercises measuring X or Y may be due to format not content. Apparently NAEP has not resolved the problem of confounding content difficulty with format difficulty (Ibid., pp. 84-85).

Final field tryouts. After exercises had been extensively revised as a result of the lay reviews, subject-matter reviews

and initial field studies, final tryouts were held prior to the selection of items for inclusion in the actual assessment (Ibid., pp. 87-97). Two general sets of procedures were used. Exercises which were to be individually administered were tried out by interviewing six persons per item. Exercises which were to be group administered were tried out by testing classroom-size sets of students. In other words, each individually administered exercise was responded to by six students, and each group exercise was responded to by about 30 persons. For the individual items, each group of six assessees included one male and one female for each of three ability levels. In terms of this tryout, if we ask whether an individually administered exercise is suitable for a girl of average ability, the answer is based on a sample of one. The same is true, of course, if we ask whether an exercise is suitable for a boy or girl of high or low ability. For group administered exercises, the classrooms were chosen on the basis of whether class members were predominantly of high or low socio-economic status.

Exercises were grouped into packages similar to those which would be used in the final assessment. Apparently not all exercises could be included, but specific information on what proportion of exercises were excluded from each assessment area is not given (Ibid., pp. 86 and 96).

The use of interviews in the tryout of individually administered exercises allowed for a relatively direct assessment of the clarity and difficulty of test items. But an indirect approach was used to assess the feasibility of the group administered exercises. The classroom teacher and a representative of the contracting agency responsible for the tryouts each completed an observation form which contained categories for such inappropriate student behaviors as apparent inattention, boredom, "cutting up", failure to follow directions and inability to cope with the task. Presumably, these behaviors may be indicators of inappropriate difficulty level or lack of clarity in some exercises. But no mention is made of whether students were interviewed to determine the reasons for the observed inappropriate behaviors (Ibid., p. 87).

NAEP drew several conclusions from the final tryout of group exercises: (1) Teachers and school administrators involved in the field trials were generally enthusiastic about the content of exercises. (2) Using a tape recorder for oral presentation of exercises only partly overcame the confounding of reading ability with knowledge of the content being tested. Some exercises are too long for students of low ability to remember the entire question. It appears that for some exercises, confounding of reading ability with knowledge of content is

exchanged for confounding of short term memory with knowledge of content. (3) The use of "I don't know" varied by ability level. High ability students tend to use it more often than do those of low ability. The number of correct responses in the low ability groups, therefore, may be spuriously large compared to the responses of more able students. Apparently, the problem of confounding format difficulty with content difficulty cannot be solved by adding "I don't know" to multiple-choice exercises. (4) Many of the problems uncovered by the field trials may be grouped and generalized to families of exercises. In the future, it may be necessary only to try out a representative sample of exercises from each type or family. (5) Inasmuch as information concerning the difficulty level of each exercise will be obtained in the final assessment, less emphasis can be given to that problem during future field trials (Ibid., pp. 89-97).

Note on clarity, difficulty and validity

The reviews and field studies used by NAEP to improve exercises have now been summarized. Our concern is whether those procedures were adequate. The first topics discussed in this section are whether NAEP's procedures were adequate in terms of the criteria of clarity and difficulty level. After that, the relevance of NAEP's procedures for content validity will be

discussed.

Clarity. This criterion was repeatedly emphasized by NAEP, as we have seen by the number of different ways in which communication problems were investigated: (1) It was one of the criteria which reviewers were asked to apply to exercises. Even the final review by the Technical Advisory Committee was concerned with possible vagueness and ambiguity of exercises (Ibid., pp. 68, 72 and 103). (2) Clarity was either the primary or secondary focus of several field studies. The unexpected high difficulty of "easy" exercises, caused the investigators in the 90 percent study to turn their attention to problems of communication and vocabulary (Ibid., p. 60). The initial feasibility studies, including the small-scale citizenship study, focused on difficulties which low achieving children have in understanding test instructions, test format and the vocabulary used in assessment exercises (Ibid., pp. 61-62 and 73). The mathematics study investigated the effect of substituting simple terms for more technical ones (Ibid., p. 76). And the final field trials of the individually administered exercises utilized interviews to uncover lack of clarity (Ibid., p. 86).

At least three improvements in exercises resulted from NAEP's concern for clarity: (1) Poor exercises were edited and re-edited using simpler and more appropriate language. Even

after the final tryouts, interview notes from the individually administered exercises were sent to the original contractors to aid in revising some exercises (Ibid., p. 97). (2) Instructions for exercises were placed on audio-tape to minimize reading difficulties. And (3) exercise formats were simplified (Ibid., p. 64).

This author is impressed with NAEP's emphasis on producing exercises which students can understand. The sheer number of reviews, field studies and revisions which were aimed at clarifying exercises is almost overwhelming. Nevertheless, NAEP seems to be operating on three questionable assumptions about how to improve exercise clarity.

The first assumption is that items which are clear to low achieving students will also be clear to more able ones (Ibid., p. 61). Although it is reasonable to assume that low achieving students will generally have more difficulty understanding instructions, formats and vocabulary than high achieving students, there may be instances in which the opposite is true. We might expect ambiguity, for instance, to present a greater problem to bright students because they may have a better grasp of the multiple meanings of some terms. A related problem is the possibility that bright students will suffer a communication handicap by being more alert to the plausibility of more than

one option being correct on multiple-choice tests. In short, NAEP policy of generally focusing on low achieving students when trying to uncover communication problems is reasonable given limited resources for field trials, but it is less satisfactory than trying out items on students with diverse abilities.

A second questionable assumption is that experts can make adequate a priori judgments about whether an exercise will be understandable to children. It is probably true that experts who have had first-hand, recent experience with children of a particular age, ability and background can make better judgments about exercises than can those who have not. But NAEP presents no evidence that the judgments of experts on review panels are an adequate substitute for trying out each exercise with students. And, although NAEP conducted several field trials which were relevant to clarity, no field trial contained all of the following features: (1) Interviews to uncover communication difficulties; (2) inclusion of all exercises which might be used in the final assessment; and (3) adequate sized samples of assessees of various ages, abilities and backgrounds.

The third questionable assumption is that it is not necessary to tryout each item (Ibid., p. 97). It is probably true that many of the problems encountered in writing clear exercises can be grouped and classified and that to uncover some of those

problems it is necessary only to sample categories of exercises rather than to try out each one. But NAEP has not demonstrated that all or even most communication problems in test items can be detected that way. It has been the experience of this author that even when a single test format is used for all exercises, and even when all items cover similar topics, problems of vagueness and ambiguity arise which are often detected only through posttest interviews with students. NAEP's recommendation that tryouts be limited to carefully selected samples of exercises is related to the assumption discussed in the preceding paragraph; it probably places too much faith in the a priori judgments of experts. To exclude exercises from the tryouts is to make a priori judgments about either the quality of those items or about their essential identity with exercises which are included. One of the useful conclusions that can be drawn from NAEP's attempt to produce exercises which meet predetermined difficulty levels is that expert judgment may not be an adequate substitute for field trials (Ibid., p. 60).

This author does not consider the above criticisms to be trivial, but they should not be used as an excuse to dismiss the extensive and worthwhile efforts made by NAEP to produce understandable exercises. The criterion of clarity is certainly important. And NAEP has substantially added to our knowledge

of how difficult it is to meet that deceptively simple standard.

Difficulty level. Three procedures were used to meet the criterion of three levels of difficulty: (1) Test writers were told to produce approximately equal numbers of exercises at each level--easy, moderate and difficult. (2) One of the tasks of the subject matter reviewers was to make a priori estimates of the difficulty of exercises. (3) Exercise difficulty was studied empirically. As we have seen, the results indicate that exercise writers seriously underestimated the difficulty of many of the so-called easy items (Ibid., pp. 57-66).

Despite the difficulty of producing easy items, the criterion of three levels of difficulty was a useful one which broke with tradition in achievement testing. Test writers usually aim for a difficulty level of about fifty percent because moderately difficult items tend to produce reliable scores. A major shortcoming of the traditional approach, however, is that it automatically excludes assessment of achievements which either nearly all or very few students can attain. In contrast, NAEP's approach was designed to provide information about the top and bottom of educational achievement, as well as the middle.

There are indications, however, that the criterion in question has become bogged down: (1) Although Tyler claimed that it is important to determine empirically whether each of

the three difficulty levels was reached, only one level was investigated in the initial field studies (Ibid., p. 53). (2) Finley and Berdie indicate that less attention should be given to pre-assessment field studies of difficulty level (1970, p. 97). And (3) it was originally planned to use the three levels of difficulty as categories for reporting NAEP findings (Merwin & Womer, 1969, p. 309; Womer, 1970, p. 11). The reports of NAEP results, however, do not appear to be organized that way. Perhaps organizing data by difficulty level did not prove to be as easy or meaningful as NAEP planners anticipated. Even so, the criterion is a good one in that it helps avoid the narrow vision of assessing only a mid-range of achievements.

Two unanticipated benefits of the criterion of three difficulty levels are worth restating. In the short run, perhaps the most important benefit is that failure to meet the criterion forced an even closer examination of the clarity of exercises. And the conclusion that content difficulty can be confounded with format difficulty may have a long-range effect on how readers interpret the kind of data gathered by NAEP.

Content validity. Although NAEP went to considerable effort to insure validity of exercises, this author believes that the project depended too heavily on a priori judgments. Dependence on those judgments seems to be rooted in NAEP's view

of the nature of content validity:

> National Assessment's one and only criterion of exercise validity is content validity . . . . If an exercise has content validity it must be an exercise that is considered to be a direct measure of some important bit of knowledge or some important skill that reflects one or more of the objectives of a subject area. In practice an exercise has content validity if it "makes sense" to an informed reader . . . (Womer, 1970, p. 9, italics added).

This author disagrees with the above definition. It can be dangerous to confuse the appearance of validity with the substance. Although face validity is perhaps the most useful approach to content validity, they should not be equated.

Before considering how NAEP might improve its procedures for determining the validity of exercises, it might be useful to consider some of the ways that exercises may fail to be valid. Most of these ways have been discussed by NAEP under various headings. (1) An exercise may be invalid if it does not correctly reflect content which it purports to measure. This is one of the problems which face validity seems best suited to handle. (2) A set of exercises may lack validity if they do not sample the total domain which they purport to measure. This problem is more complicated than simply examining exercises to see if they make sense. The analysis of a domain into categories and the tabulation of exercises which fit those categories seems called for, but the heart of the matter is still face

validity. (3) An exercise may be invalid if the assessees do not understand the question or task. This is NAEP's criterion of clarity. Something similar to face validity, the a priori judgments of experts, may be useful, but the direct assessment of clarity is more appropriate. Interviews of assessees seem called for. (4) An exercise may be invalid if it means different things to different persons. This is a different problem than clarity. It is possible that students from different ethnic or social class backgrounds will literally respond to different issues when confronted with identical exercises, especially exercises having to do with values. Students bring meaning to test items; they do more than extract meaning from them. In detecting these differences, face validity cannot adequately replace interviewing assessees. (5) An exercise may be invalid if it confounds measurement of the intended achievement with measurement of an extraneous variable. For instance, the mathematics study and the choices study indicate that some exercises confound format difficulty with content difficulty.

Obviously, NAEP's concern for clarity and difficulty-level are pertinent to the above comments. It is also obvious that NAEP's reviews and field studies amount to a serious attempt to improve exercise validity. But three suggestions might be helpful:

(1) NAEP should make a greater effort to determine what meaning social studies and citizenship exercises have for different students. Not only should a greater effort be made to determine whether assessees understand the task, but also the basis for their response. It might be that the directionality of some exercises is too narrowly restricted by white or middle-class viewpoints (Citizenship: Group Results B, 1972, p. 54). More extensive use of interviews in field studies seems warranted.

(2) This author has found no mention of item analysis* or factor analysis in NAEP reports. It is assumed that factor analysis is considered inappropriate for single-item exercises, especially since different students respond to different packages of exercises. Although these procedures may not be appropriate in the actual assessments, they might be useful during field trials. Item analysis may be useful for spotting odd patterns of response which might indicate vague or ambiguous exercises which slipped past the reviews. Factor analysis might be a useful procedure for item reduction. And, inasmuch as some NAEP reports cluster exercises when discussing findings, factor

---

*It could be argued that NAEP's reporting procedure amounts to item analysis. What role, if any, the analysis of response patterns played in the development of exercises is not noted in NAEP publications.

analysis in pilot studies may shed light on whether such clusters make sense; it may provide information on whether a cluster of exercises seems to measure the same objective. If this author had to choose between factor analysis and face validity, he would opt for the latter, but that choice may not be necessary.

(3) NAEP refrains from comparing results to different exercises, unless the compared exercises use identical formats and have equally difficult distractors (Ibid., p. 39). Readers, however, will be tempted to compare the percentage of correct response across exercises. NAEP should warn that such comparisons may be misleading.

The thrust of the second and third points above is that NAEP may be relying too heavily on a priori judgments of exercise validity. They should not be taken to mean that NAEP did not make extensive and serious efforts to produce valid exercises.

Reliability

NAEP's approach to reliability is related to several innovations in assessing achievement. Most of those innovations were adopted because some assumptions underlying standard approaches to achievement testing were rejected as inappropriate

to the purposes of NAEP. One of those assumptions is that good achievement tests will differentiate among individuals who differ in knowledge of the subject being tested. That assumption is appropriate for some purposes. If we are testing for the purpose of assigning grades to students we want our instrument to reliably differentiate among excellent, average and poor students. But the purposes of NAEP are different. What NAEP sought to determine was what Americans in general know of certain topics at certain ages, not what any individual knows.

The high reliability needed to differentiate among individuals is usually gained by combining a number of test items into a single instrument; for instance, a fifty-item multiple-choice test in American History. The student's response to a total set of items is more reliable than his or her response to any one question. This approach, however, is not necessary if the intent is to assess the achievement of large groups rather than individuals. Reliable estimates of group performance can be obtained by summing the responses of large numbers of people to a single item, just as reliable estimates of individual performance can be obtained by summing the responses of one person across several items.

Besides being unnecessary, there are positive disadvantages to using standard types of achievement tests to assess the

performance of large groups. One disadvantage is that test scores mask important information. A score of 40 out of 50 items does not reveal the specific weaknesses or strengths in a student's knowledge. In fact, two students with identical scores may have very different strengths and weaknesses in knowledge of the subject tested. By analogy, had NAEP used a single standard-type test in citizenship or social studies, very little information about national achievement levels would have been revealed.

Therefore, as noted previously, NAEP did not develop achievement tests in the usual sense. Instead, a large number of individual exercises were constructed, each one testing a specific belief, knowledge, or other type of competency. These exercises are analogous to single-item tests. Results are reported as that proportion of the sample which correctly responds to an exercise.

This approach allowed for another innovation; not everyone in the sample responded to the same set of exercises. Instead, exercises were grouped into different "packages" which were randomly assigned to persons in the sample. In this way it was possible to obtain an adequate number of responses to far more exercises than could have been included in any single achievement test. In other words, the domain of achievement which was

tested was much larger than would have been possible had standard achievement tests been employed. This benefit outweighs the relatively slight disadvantage of having to select a larger sample than would have been necessary had each person taken the same test.

Because of the unusual features of NAEP exercises, standard estimates of reliability are inappropriate. Readers who are accustomed to seeing reliability reported as coefficients, such as .84 or .91, may be puzzled by the lack of such statements in NAEP reports. Types of reliability, however, which yield coefficients are based on those assumptions about achievement testing which NAEP rejected. An alternative way of estimating reliability is to compute standard errors. This approach will be unfamiliar to many readers. Nevertheless they can depend on NAEP reports to be cautious in the narrative descriptions of findings. The language used to discuss findings is carefully chosen to reflect the amount of trust that can be placed on their reliability. Readers who wish a more technical discussion of this problem should consult Appendix F of NAEP Preliminary Report 9: Citizenship, Group Results, 1972.

Sample

A serious shortcoming of most educational research is the lack of money and manpower to select and test large samples which represent national, regional or other significant populations. Most often, samples are composed of local volunteers such as a few social studies classes in schools that are willing to cooperate. Samples are also frequently composed of captive audiences such as students enrolled in courses taught by the person doing the research. Results from such samples may shed light on important educational problems, but doubts about the similarity between the students tested and other populations cannot be dismissed.

In contrast, NAEP studies employ careful and thorough sampling procedures. Readers of NAEP reports can be reasonably certain, for instance, that the NAEP sample of nine-year-old children is similar to the national population of the same age. Similarly, NAEP's sample of 13-year-old children who live in the Northeast is representative of youth of the same age who live in that section of the country.

Of course, no sampling procedure is without error, but if the sample is drawn properly error can be minimized and estimated. The best sampling procedures involve random selection, and large random samples contain less sampling error than

small ones. NAEP obtained large samples of each of the sub-populations for which they intended to report results, and the samples were based on a modified random selection.

Selecting random national samples is difficult. Simple random sampling requires that the name of each person in the target population be placed on a list; for example, each nine-year-old child in the United States. After the list is complete, the names to be included in the sample are drawn at random. It is impractical, if not impossible, however, to list every person in the United States who falls within a given age-group. And even if such a listing were possible, it would be unnecessarily time consuming and expensive.

Instead of simple random sampling, NAEP combined randomization with multi-stage cluster sampling. In the sense that it is used here, clusters refer to groups of people. It makes sense to think of students as clustered or grouped by geographic regions and by political units such as state, county, school district and so on. The term "multi-stage" means that the sample is drawn in several stages or steps. In the early stages of cluster sampling, groups are randomly selected; for instance, so many schools, in so many counties, in so many states. In the final stage, individual students are selected at random from the schools selected in the previous stage. If properly

conducted, these procedures produce representative samples. The obvious advantage over simple random sampling is that it is much easier to construct lists of relevant clusters than to list every person in the nation.

Some sense of the sampling tasks faced by NAEP can be gained by noting that samples were drawn which were representative of four age levels, four national regions, and communities of various types and sizes. At some age levels, special precautions were taken to include adequate numbers of persons of low socio-educational status. Furthermore, testing of all of one age-group, the young adults, plus that portion of the 17-year-old group which was no longer enrolled in school, was conducted in the homes of the persons included in the sample. This called for additional adjustments in the way samples were drawn.

NAEP has drawn attention to several limitations in the samples. NAEP did not intend to report findings on a state by state basis. Therefore, sub-samples were not drawn to be representative of various states. Persons interested in knowing how students in their state compare to students from other geographic or political units cannot obtain that information from NAEP.

A second limitation is that the original sampling design did not include race as one of its strata. Although a large

enough sample of black assessees was obtained to compare the performance of blacks to non-blacks, adequate samples of minorities other than black youth and adults were not obtained.

A third limitation is that the defined sample and the obtained sample are not identical. Each person in each sample had the right to refuse to take part in the assessment. This was not a serious problem in the schools; more than nine out of every ten studerts agreed to respond to the exercises. But, during the first assessment year, only 60 percent of the young adults and out-of-school 17-year-olds who were contacted agreed to cooperate. Readers of NAEP reports for the first assessment year, which includes citizenship, should not place as much faith in the representativeness of the out-of-school sample as they do in the in-school sample (Womer, 1970, p. 34).

Readers interested in a fairly technical explanation of NAEP samples should consult Appendix C of NAEP Report I: Science, National Results, 1970. Some changes have been made in the way samples were drawn in subsequent assessments, but the basic approach is similar. A more readable description of sampling procedures can be found in Frank B. Womer, What is National Assessment, pages 22-28. The overall impression left by these reports is that ability to handle technical problems is one of the strengths of NAEP.

The impressive technical quality of NAEP is probably due to the expertise of professionals associated with the project. Ralph Tyler was instrumental in founding NAEP. The technical advisory committee has included such noted scholars as Robert P. Abelson, Lee J. Cronbach, Lyle V. Jones and John W. Tukey. Personal interviews with some members of the NAEP staff who are responsible for sampling, data collection and analysis indicate that they are competent professionals. Furthermore, NAEP has frequently contracted with private agencies when expertise was not available within the project. The wealth of talent that has been available to NAEP stands in marked contrast to the one or two-person staffs which are characteristic of most educational research.

Data Collection

Data collection for a project as large as NAEP must be carefully planned and coordinated. It requires cooperation between the schools and NAEP staff. And it requires honesty and competence on the part of those collecting data.

Apparently, NAEP's data collection procedures were carefully planned and coordinated, and they were successful in obtaining the cooperation of the schools (Womer, 1970, pp. 28-35). This author has no reason to believe that NAEP field staff were

either incompetent or dishonest, but the opportunity for some of the problems associated with hired-hand research seem to have been present. In fact, they may be unavoidable in large projects.

Hired-hand research refers to projects which utilize lower-level staff members for data collection (Roth, 1966). Hired-hands may be graduate students, housewives or others who need part-time or short-term employment. For a number of reasons, hired-hands sometimes fake data. Their job may be boring. Patterns of response may appear to be predictable, so the hired-hand invents data which he believes fit the patterns which would have emerged from actual data. Data collectors may feel threatened when venturing into certain neighborhoods or when knocking on certain doors, so data is faked for those locations. Hired-hands may fail to see the importance of a study. The information gathered may appear pointless, so they see no need for care and accuracy.

How problems associated with hired-hand research might occur in NAEP is evident in the procedures used to collect data. Data were collected in two types of settings--schools and homes. And they were collected in two ways--group administered exercises and individual interviews. Two types of personnel were used--district supervisors and exercise administrators. There

were twenty seven district supervisors, each responsible for a geographic area. Duties of the district supervisors included making arrangements with the schools involved, hiring, training and supervising the exercise administrators. Exercise administrators were " . . . recruited from lists of substitute teachers, from among college trained housewives, from graduate students . . . and from other sources of competent adults (Womer, 1970, p. 30).

In this author's judgment, the household interviews of adults and out-of-school 17-year-olds during the first assessment year was the most likely place for problems of hired-hand research to occur. The low rate of cooperation in the out-of-school assessment for that year may indicate that the work of the exercise administrators was more threatening or frustrating when going from door to door than when giving tests in schools. It may also indicate that the exercise administrators were not competent to conduct household interviews. A third possibility, is that some of the potential assessees were offended by the race, age or apparent social class of the interviewer. And, of course, it just might be that 40 percent of the people didn't want to be messed with, by anyone. NAEP's characteristic care in technical matters is evidenced by the fact that a special quality check was conducted to determine whether bias was

introduced into the data as a result of the low percent of cooperation in the initial out-of-school study (Ibid., p. 34).

Other quality checks are also utilized, at least in the more recent assessments. District supervisors revisit a sample of the homes visited by the exercise administrators. A mail-validation procedure is used to check on the quality of the interview data. And computer checks are made to help spot patterns of response which deviate from normal. The out-of-school cooperation rate for the second-year assessment jumped from about 60% to 75% for adults and 97% for 17-year-olds (Reading and Literature: General Information Yearbook). Whether this is due to improved quality checks by NAEP or to the fact that in-home assessees are offered money to cooperate is not known.

Data Analysis and Reporting

From the standpoint of competence and care, data analysis is one of the most impressive aspects of NAEP. For the most part, the procedures used are simple and readers of NAEP reports are given adequate warning as to what types of conclusions can be legitimately drawn from the data. Many of the results are reported as proportions of correct response to an exercise. When comparisons are made between groups, such as by sex, age, race or region, readers are informed as to whether the differences

are reliable. And most important, readers are frequently warned not to draw conclusions concerning causation. Readers are even warned to be careful concerning the meaning that is attached to labels given to groups (Ibid., p. v).

Perhaps the most complicated analytic procedure used by NAEP is balancing, which is an attempt to remove the masking of variables. The following is an example of masking: When comparisons are made between the performance of inner-city and suburban children, differences in achievement may be due not to the location of the child's residence but to the socio-educational status of his family. Although balancing seems to offer some promise of weeding out these confounding variables, it is far from a perfect solution, as NAEP repeatedly reminds us (Ibid., pp. 49-51).

One of the few "goofs" in the analysis and reporting of data is important, but non-technical. We wonder how often NAEP staff have had the following statement called to their attention: "The male-female difference is <u>positive</u> if the males have a higher percentage of success than females. The male-female difference is <u>negative</u> if the females have a higher percentage of success than the males (Ibid., p. 37)." This error in judgment is consistent with this author's impression that NAEP is less vulnerable to criticism on technical than on

non-technical matters.

How best to present its findings is one of the perplexing non-technical issues confronting NAEP. At one point it appeared that NAEP intended to organize findings around the difficulty of exercises, but that approach apparently was dropped. Another approach is to simply report each finding for each exercise, but that approach tends to overwhelm the reader. An alternative is to report findings by clusters of exercises called themes (Ibid., p. i-ii). It appears that the social studies report titled Political Knowledge and Attitudes is organized that way, and it seems to be more readable than the previous citizenship reports. A major difficulty remains, however, which seems to be characteristic of all NAEP reports of data: After having read a report, it is difficult to pull together a coherent, meaningful set of conclusions. This author finds himself either repeating the numerous specific findings or, going to the other extreme, drawing the most general conclusions. Among those general conclusions are such statements as: The Southeast does not do as well as the Northeast; Inner-city students do not do as well as suburban children; Blacks generally do not do as well as whites. The limited utility of such conclusions is obvious.

Perhaps the problem is that NAEP is hesitant to interpret its own data. If so, it may be useful to have outside

critics attempt the task of interpretation, assuming that meaningful and useful conclusions can be legitimately drawn from the type of information which NAEP has gathered.

Conclusion

The general tone of a critique, positive or negative, is probably as much a reflection of the personality and competence of the critic as of the quality of the object criticized. Readers of this report should keep that in mind when noting that there is a marked difference in the amount of attention given to different aspects of NAEP. Furthermore, in this report, the amount of discussion of any topic, such as data analysis or exercise development, is related to the author's judgment concerning its vulnerability to negative criticism. That some topics have received more criticism than others, however, should not be taken to mean that the author believes that the overall quality of any aspect of NAEP is more negative than positive. The opposite is true.

Despite the several criticisms and suggestions in this report, it is the overall impression of this writer that NAEP used reasonable procedures. The technical aspects of the project appear to be sound. Most of the innovations, such as using laymen to review exercises, appear to be useful. Although a project

as massive as NAEP is bound to run into difficulties, even some failures, there are few, if any, research efforts in education of this scope and quality.

References

Finley, C. J., & Berdie, F. S. The National Assessment Approach to Exercise Development, Denver, National Assessment of Educational Progress, 1970.

Hunkins, F. P. Validity of Social Studies and Citizenship Exercises, Task II, Final Report, University of Washington (mimeo), 1973.

National Assessment of Educational Progress, Citizenship, National Results, 1970, Denver, National Assessment of Educational Progress, 1970.

National Assessment of Educational Progress, Summary of Discussion, Conference for Lay Review of Sample Exercises, Chicago, May 10-12, 1966 (mimeo, no publication site or date given).

National Assessment of Educational Progress, Conferences on the Development of Instrumentation for Assessment of the Progress of Education in the United States, June 15, 1965 - June 30, 1967 (mimeo, no publication site or date given).

National Assessment of Educational Progress, Summary of Development of Citizenship -- 1965-1969 (mimeo, no publication site or date given).

National Assessment of Educational Progress, Item Development: Social Studies (mimeo, no publication site or date given).

National Assessment of Educational Progress, Social Studies Review, NAEP Offices, Ann Arbor, Michigan, August 22-23, 1970 (mimeo, no publication site or date given).

National Assessment of Educational Progress, Subject Matter Review, October 8-11, 1970, Campus Inn, Ann Arbor, Michigan (mimeo, no publication site or date given).

National Assessment of Educational Progress, Social Studies Writing Conference, Ann Arbor, Michigan - Thursday, April 23-Sunday, April 26, 1970 (mimeo, no publication site or date given).

National Assessment of Educational Progress, Citizenship, Group Results B, 1972, Denver, National Assessment of Educational Progress, 1972.

National Assessment of Educational Progress, Science, National Results, 1970, Denver, National Assessment of Educational Progress, 1970.

National Assessment of Educational Progress, Political Knowledge and Attitudes, 1971-1972, Denver, National Assessment of Educational Progress, 1973.

Merwin, J. C., & Womer, F. B. Evaluation in Assessing the Progress of Education to Provide Bases of Public Understanding and Public Policy, Educational Evaluation: New Roles, New Means (Ralph W. Tyler, Ed.), Chicago, The University of Chicago Press, 1969.

Roth, J. Hired Hand Research, American Sociologist 1 (August 1966), pp. 190-96.

Womer, F. B., What is National Assessment? Denver, National Assessment of Educational Progress, 1970.